



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ
ΤΟΜΕΑΣ ΣΥΣΤΗΜΑΤΩΝ ΜΕΤΑΔΟΣΗΣ ΠΛΗΡΟΦΟΡΙΑΣ & ΤΕΧΝΟΛΟΓΙΑΣ ΥΛΙΚΩΝ

Χρήση δικτύων παραγωγικής μάθησης και διαδικτύου των πραγμάτων για την δημιουργία συστάσεων διατροφής

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

της

ΚΑΤΣΑΜΠΟΥΛΑ ΧΡΙΣΤΙΝΑΣ ΣΟΦΙΑΣ

Επιβλέπων : Κωνσταντίνα Νικήτα
Καθηγήτρια Ε.Μ.Π.

Αθήνα, Ιούνιος 2022



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ
ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ
ΤΟΜΕΑΣ ΣΥΣΤΗΜΑΤΩΝ ΜΕΤΑΔΟΣΗΣ ΠΛΗΡΟΦΟΡΙΑΣ & ΤΕΧΝΟΛΟΓΙΑΣ
ΥΛΙΚΩΝ

**Χρήση δικτύων παραγωγικής μάθησης και διαδικτύου
των πραγμάτων για την δημιουργία συστάσεων
διατροφής**

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

της

ΚΑΤΣΑΜΠΟΥΛΑ ΧΡΙΣΤΙΝΑΣ ΣΟΦΙΑΣ

Επιβλέπων : Κωνσταντίνα Νικήτα
Καθηγήτρια Ε.Μ.Π.

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 29^η Ιουνίου 2022.

(Υπογραφή)

.....
Κωνσταντίνα Νικήτα
Καθηγήτρια Ε.Μ.Π.

(Υπογραφή)

.....
Ανδρέας-Γεώργιος Σταφυλοπάτης
Καθηγητής Ε.Μ.Π.

(Υπογραφή)

.....
Γεώργιος Στάμου
Αναπληρωτής Καθηγητής Ε.Μ.Π.

Αθήνα, Ιούνιος 2022

(Υπογραφή)

.....

ΚΑΤΣΑΜΠΟΥΛΑ ΧΡΙΣΤΙΝΑ ΣΟΦΙΑ

Διπλωματούχος Ηλεκτρολόγος Μηχανικός και Μηχανικός Υπολογιστών Ε.Μ.Π.

Copyright © Χριστίνα Σοφία Κατσάμπουλα, 2022.

Με επιφύλαξη παντός δικαιώματος. All rights reserved.

Απαγορεύεται η αντιγραφή, αποθήκευση και διανομή της παρούσας εργασίας, εξ' ολοκλήρου ή τμήματος αυτής, για εμπορικό σκοπό. Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα. Ερωτήματα που αφορούν τη χρήση της εργασίας για κερδοσκοπικό σκοπό πρέπει να απευθύνονται προς τον συγγραφέα.

Οι απόψεις και τα συμπεράσματα που περιέχονται σε αυτό το έγγραφο εκφράζουν τον συγγραφέα και δεν πρέπει να ερμηνευθεί ότι αντιπροσωπεύουν τις επίσημες θέσεις του Εθνικού Μετσόβιου Πολυτεχνείου.

Περίληψη

Μέρα με την μέρα η τεχνολογία εισβάλλει όλο και περισσότερο στην καθημερινότητά μας, αλλάζοντας τις καθημερινές μας συνήθειες. Οι συσκευές που χρησιμοποιούμε τείνουν να γίνονται ολοένα και πιο έξυπνες. Ο κλάδος των smart things έχει αρχίσει να αναπτύσσεται με γοργούς ρυθμούς, και όλο και περισσότερες συσκευές αποκτούν συνδεσιμότητα στο διαδίκτυο. Με αυτή την σκέψη, στο πλαίσιο της παρούσας διπλωματικής εργασίας, θελήσαμε να δημιουργήσουμε μια προσέγγιση έξυπνου ψυγείου, το οποίο με βάση το περιεχόμενό του μπορεί να προτείνει στον χρήστη γεύματα που οδηγούν σε μια ισορροπημένη διατροφή. Αρχικά ο χρήστης μέσω φορητής κάμερας φωτογραφίζει τα περιεχόμενα του ψυγείου του. Με τη βοήθεια του εργαλείου Google Vision API, αναγνωρίζονται τα περιεχόμενα της φωτογραφίας και αποδίδονται ετικέτες σε αυτά. Με βάση το περιεχόμενο που αναγνωρίστηκε, επιλέγεται κατάλληλο σύνολο δεδομένων που χρησιμοποιείται για την εκπαίδευση ενός παραγωγικού ανταγωνιστικού δικτύου. Συγκεκριμένα χρησιμοποιήθηκε ένα DCGAN (Image-Generator) που παίρνοντας σαν είσοδο σύνολα εικόνων καταφέρνει να δημιουργεί συνθετικές εικόνες γευμάτων που μπορούν να παρασκευαστούν από τα υλικά που αναγνωρίστηκαν στην φωτογραφία των περιεχομένων του ψυγείου. Για την αξιολόγηση των παραγόμενων εικόνων και ειδικότερα της απόδοσης του δικτύου, χρησιμοποιήθηκε ένα συνελκτικό νευρωνικό δίκτυο, το οποίο εκπαιδεύτηκε σε δύο κλάσεις, ώστε να αναγνωρίζει δύο κατηγορίες τροφής. Το πείραμα εκτελέστηκε δύο φορές με διαφορετικό συνδυασμό κλάσεων. Ειδικότερα στο πρώτο πείραμα, το συνελκτικό νευρωνικό δίκτυο εκπαιδεύτηκε με πραγματικές εικόνες ώστε να αναγνωρίζει ψωμί και ζυμαρικά. Στη συνέχεια, κλήθηκε να αναγνωρίσει συνθετικές εικόνες ψωμιού και ζυμαρικών, οι οποίες παράγονται από το παραγωγικό ανταγωνιστικό δίκτυο. Το συνελκτικό δίκτυο κατάφερε να αναγνωρίσει τις παραγόμενες εικόνες ζυμαρικών και ψωμιού με επιτυχία 71%. Στο δεύτερο πείραμα, το δίκτυο εκπαιδεύτηκε να αναγνωρίζει εικόνες ψωμιού και λαχανικών και στη συνέχεια κατάφερε να ταξινομήσει συνθετικές εικόνες ψωμιού και λαχανικών με επιτυχία 87%. Τα αποτελέσματα αυτά δείχνουν ότι το παραγωγικό ανταγωνιστικό δίκτυο μπορεί να δημιουργήσει συνθετικές εικόνες τροφών, τις οποίες το συνελκτικό νευρωνικό δίκτυο καταφέρνει να αναγνωρίσει ορθά στην πλειονότητά τους, έπειτα από εκπαίδευση με πραγματικές εικόνες.

Λέξεις Κλειδιά: Διαδίκτυο των πραγμάτων, Έξυπνες συσκευές, Έξυπνο ψυγείο, Όραση υπολογιστών, Τεχνητή νοημοσύνη, Μηχανική μάθηση, Συνελκτικά Νευρωνικά Δίκτυα, Παραγωγικά Ανταγωνιστικά Δίκτυα, Confusion Matrix, Google Vision API, Graphical User Interface, Image-Generator

Abstract

Technology is invading more and more into our daily lives, changing our daily habits. The devices that we use are becoming smarter. The field of smart things has started to grow rapidly and a large amount of devices are gaining access to the internet. Having that in mind, we had the ambition to create an approach of smart fridge able to propose healthy and balanced food meals, based on refrigerator's contents. Initially, the user takes a photo of their refrigerator's content which is analyzed through the Google Vision API, a tool provided by the Google Platform. The Google Vision API recognizes the objects in the photo and provides labels for the available food ingredients. Using these labels, an appropriate training set is selected and used in order to train a generative adversarial network. Within the framework of this thesis, the Image-Generator was used, a DCGAN that is fed with image datasets and creates synthesized food images which represent feasible cooking options for a healthy and balanced diet, based on the available ingredients. In order to evaluate the performance of the DCGAN, we used a two-class convolutional neural network (CNN) which was trained with actual images to recognize two categories of food. The evaluation procedure was carried out for two different combinations of food classes, involving recognition of (i) bread vs pasta images and (ii) bread vs vegetables images. After proper training with real images, the CNN model was able to successfully classify 71% of the generated bread and pasta images and 87% of the generated bread and vegetables images. These results show that the generative adversarial network is able to create synthetic images which can be correctly classified by the CNN.

Keywords: Internet of things, Smart Things, Smart Fridge, Computer Vision, Artificial Intelligence, Machine Learning, Convolutional Neural Networks, Generative Adversarial Networks, Confusion Matrix, Google Vision API, Graphical User Interface, Image-Generator

Ευχαριστίες

Πρώτα απ' όλα θα ήθελα να ευχαριστήσω την επιβλέπουσα καθηγήτρια μου, κα. Νικήτα, όπως και τους συνεργάτες της, την μεταδιδακτορικό ερευνήτρια Καλλιόπη Δαλακλειδίη και τον υποψήφιο διδάκτορα Θεοφάνη Γανιτίδη, για την συνεργασία, την καθοδήγηση και την υπομονή τους καθ' όλη την διάρκεια της διπλωματικής μου εργασίας. Ήταν πάντα πρόθυμοι να μου απαντήσουν οποιαδήποτε απορία και αν είχα, κάτι το οποίο ήταν πολύτιμο.

Επίσης θα ήθελα να ευχαριστήσω τους γονείς μου και τα αδέρφια μου που πάντα με στηρίζουν και με εμπυχώνουν σε όλες τις δύσκολες στιγμές. Τίποτα δεν θα ήταν το ίδιο χωρίς την δική τους παρουσία στην ζωή μου και ειδικότερα σε αυτό το μονοπάτι της τριτοβάθμιας εκπαίδευσης.

Ένα πολύ μεγάλο ευχαριστώ θέλω να πω και στον Μίλτο που ήταν πάντα βράχος δίπλα μου, με πίστευε και δεν έπαψε ποτέ να μου δίνει κουράγιο και δύναμη να συνεχίσω. Επιπλέον ευχαριστώ πολύ τους αγαπημένους μου φίλους και κυρίως τις φίλες μου που μου συμπαραστάθηκαν όλα αυτά τα χρόνια, και αφουγκράστηκαν τις έγνοιες μου και τα άγχη μου.

Τέλος θέλω μέσα από την καρδιά μου να ευχαριστήσω όλους μου τους φίλους και συμφοιτητές που μοιράστηκαν μαζί μου αυτό το ταξίδι. Άπειρες αναμνήσεις. Άπειρο διάβασμα. Άπειρη γνώση. Μοιραστήκαμε όμορφες, ευχάριστες, αστείες αλλά και στενάχωρες στιγμές. Ήμασταν όμως όλοι μαζί και όλα έμοιαζαν ευκολότερα. Μάθαμε τι θα πει συνεργασία, αλληλοκατανόηση, αλληλοβοήθεια. Θέλω να τους ευχαριστήσω όλους έναν έναν ξεχωριστά.

Χριστίνα Σοφία Κατσάμπουλα
29/06/2022

Πίνακας περιεχομένων

Κεφάλαιο 1: Εισαγωγή	21
1.1 Ιστορική αναδρομή	21
1.2 Σκοπός της παρούσας εργασίας	21
1.3 Σχετική βιβλιογραφία.....	22
1.3.1 Cook-GAN.....	22
1.3.2 Gan-stronomy: Generative cooking with conditional DCGANs.....	23
1.3.3 A Food Dish Image Generation Framework Based on Progressive Growing GANs	24
1.3.4 Food Image Generation using A Large Amount of Food Images with Conditional GAN: RamenGAN and RecipeGAN	25
1.3.4.1 RamenGAN.....	25
1.3.4.2 RecipeGAN.....	26
Κεφάλαιο 2: Θεωρητικό υπόβαθρο	28
2.1 Διαδίκτυο των πραγμάτων	28
2.1.1 Αρχιτεκτονική του διαδικτύου των πραγμάτων.....	29
2.1.1.1 To Perception Layer	29
2.1.1.2 To Network Layer	29
2.1.1.3 To Application Layer	30
2.2 Έξυπνες συσκευές.....	30
2.2.1 Διαχωρισμός των έξυπνων συσκευών.....	31
2.3 Όραση υπολογιστών	32
2.4 Τεχνητή Νοημοσύνη	32
2.5 Μηχανική Μάθηση	33
2.5.1 Επιβλεπόμενη Μάθηση.....	33
2.5.2 Εκμάθηση χωρίς επίβλεψη.....	34
2.6 Βαθιά μάθηση	34
2.7 Τεχνητά Νευρωνικά Δίκτυα	34
2.8 Συνελκτικά Νευρωνικά δίκτυα	35
2.8.1 Συνελκτικό επίπεδο	35
2.8.2 Στρώμα συγκέντρωσης	36
2.8.3 Πλήρως συνδεδεμένο επίπεδο	36
2.8.4 Συνελκτικό νευρωνικό δίκτυο για το σύνολο δεδομένων MNIST	36
2.9 Παραγωγικά Ανταγωνιστικά Δίκτυα	37
2.9.1 Λειτουργία Παραγωγικών Ανταγωνιστικών Δικτύων	38
2.9.2 Αρχιτεκτονικές Παραγωγικών Ανταγωνιστικών Δικτύων	39
2.9.2.1 Πλήρως συνδεδεμένα Παραγωγικά Ανταγωνιστικά Δίκτυα	39

2.9.2.2 Συνελκτικά Παραγωγικά Ανταγωνιστικά Δίκτυα	39
2.9.2.3 Υπό όρους Παραγωγικά Ανταγωνιστικά Δίκτυα.....	39
2.9.2.4 GAN με μοντέλα συμπερασμάτων	39
2.9.2.5 Adversarial Autoencoders	39
2.10 Υπερπροσαρμογή.....	40
2.11 Αξιολόγηση των Παραγωγικών Ανταγωνιστικών Δικτύων	40
2.11.1 GAN- train Scores.....	40
2.11.2 GAN- test Scores	41
2.11.3 Μήτρα σύγχυσης.....	41
2.11.4 True Positive Rate	41
2.11.5 True Negative Rate.....	42
2.11.6 False Negative Rate.....	42
2.11.7 False Positive Rate	42
2.11.8 Accuracy.....	43
2.11.9 Balanced Accuracy	43
2.11.10 Positive Predictive value.....	43
2.11.11 Negative Predictive value.....	43
2.11.12 F1 score	44
Κεφάλαιο 3: Πειραματική Διαδικασία.....	45
3.1 Graphical User Interface	45
3.2 Google Cloud's Vision API.....	46
3.3 Επιλογή ετικετών	50
3.4 Δημιουργία εικόνας με την βοήθεια Παραγωγικών ανταγωνιστικών δικτύων	50
3.4.1 Επιλογή συνόλου δεδομένων για την εκπαίδευση του παραγωγικού ανταγωνιστικού δικτύου	51
3.4.2 Παραγωγικό ανταγωνιστικό δίκτυο: Image-Generator.....	53
3.4.3 Αρχιτεκτονική Image-Generator	55
3.4.4 Προσδιορισμός κατάλληλων παραμέτρων για το Image-Generator και προσαρμογή του δικτύου αυτού για την παρούσα εργασία.....	56
3.5 Σχολιασμός αποτελέσματος.....	60
3.6 Παραγόμενες εικόνες από το Image-Generator	60
3.7 Σχολιασμός παραγόμενων εικόνων.....	64
Κεφάλαιο 4: Αξιολόγηση παραγόμενων εικόνων	66
4.1 Αξιολόγηση των παραγόμενων εικόνων για τις κατηγορίες ψωμί και μακαρόνια	66
4.1.1 Εκπαίδευση του συνελκτικού νευρωνικού δικτύου.....	66

4.1.2 Αξιολόγηση των παραγόμενων εικόνων για τις κατηγορίες ψωμί και μακαρόνια με την βοήθεια του εκπαιδευμένου συνελκτικού νευρωνικού δικτύου.	72
4.2 Αξιολόγηση των παραγόμενων εικόνων για τις κατηγορίες ψωμί και λαχανικών.	74
4.2.1 Εκπαίδευση του συνελκτικού νευρωνικού δικτύου.....	74
4.2.2 Αξιολόγηση των παραγόμενων εικόνων για τις κατηγορίες ψωμί και λαχανικών με την βοήθεια του εκπαιδευμένου συνελκτικού δικτύου.....	80
4.3 Σύγκριση αποτελεσμάτων για τα δύο πειράματα.	81
Κεφάλαιο 5: Συμπεράσματα και Μελλοντικές Προεκτάσεις της Εργασίας.....	83
Βιβλιογραφία	84

Κατάλογος σχημάτων

Εικόνα 1. 1: Το μοντέλο που απεικονίζει την αξιολόγηση των εικόνων και αν αυτές είναι ρεαλιστικές ή όχι (discriminator) [10].	23
Εικόνα 1. 2: Αποτελέσματα του Gan-stronomy [11].	24
Εικόνα 1. 3: Αναπαράσταση της αρχιτεκτονικής [12].	25
Εικόνα 1. 4: Απεικονίζονται τα τρία νευρωνικά δίκτυα, δηλαδή ο generator και οι δύο discriminators [16].	26
Εικόνα 1. 5: Αριστερά: Αξιολόγηση εικόνας που έχει προέλθει από απλό GAN. Η πάνω είναι η παραγόμενη εικόνα και η κάτω είναι η εικόνα μετά την αναζήτηση με “φίλτρο” την παραγόμενη εικόνα. Δεξιά: Αξιολόγηση εικόνας που έχει προέλθει από WGAN-GP. Η πάνω είναι η παραγόμενη εικόνα και η κάτω είναι η εικόνα μετά την αναζήτηση με “φίλτρο” την παραγόμενη εικόνα [16].	27
Εικόνα 2. 1: Η εξέλιξη του Internet of Things [1].	29
Εικόνα 2. 2: Τα τρία επίπεδα της αρχιτεκτονικής του Internet of Things [21].	30
Εικόνα 2. 3 : Αριστερά απεικονίζεται ένα τεχνητό νευρωνικό δίκτυο 2 επιπέδων το οποίο αποτελείται από ένα επίπεδο εισόδου τριών εισόδων, από ένα κρυφό επίπεδο τεσσάρων νευρώνων και από ένα επίπεδο εξόδου δύο νευρώνων. Δεξιά απεικονίζεται ένα τεχνητό νευρωνικό δίκτυο 3 επιπέδων το οποίο αποτελείται από ένα επίπεδο εισόδου τριών εισόδων, από δύο κρυφά επίπεδα τεσσάρων νευρώνων το καθένα και από ένα επίπεδο εξόδου ενός νευρώνα. Αξίζει να παρατηρηθεί ότι και στα δύο παραδείγματα νευρωνικών δικτύων υπάρχουν συνάξεις, δηλαδή συνδέσεις, ανάμεσα στους νευρώνες που βρίσκονται σε διαφορετικό επίπεδο και όχι στο ίδιο. Η εικόνα πάρηκε από [39].	35
Εικόνα 2. 4: Μια αναπαράσταση της συνέλιξης [40].	36
Εικόνα 2. 5: Μια απλή αρχιτεκτονική ενός συνελκτικού νευρωνικού δικτύου [41].	37
Εικόνα 2. 6: Μια αρχιτεκτονική ενός συνελκτικού νευρωνικού δικτύου με πολυπλοκότερη δομή [41].	37
Εικόνα 2. 7: Περιγράφεται σχηματικά η λειτουργία των GANs.	38
Εικόνα 3. 1: Απεικονίζεται το GUI και βλέπουμε το κουμπί που εάν πατηθεί θα ανοίξει και θα ενεργοποιηθεί η κάμερα του υπολογιστή.	45
Εικόνα 3. 2: Απεικονίζεται το παράθυρο με την κάμερα του υπολογιστή που εάν πατηθεί το “spacebar” θα ληφθεί η φωτογραφία.	46
Εικόνα 3. 3: Απεικονίζεται το GUI και βλέπουμε το κουμπί που εάν πατηθεί θα ανοίξει και θα ενεργοποιηθεί η κάμερα του υπολογιστή αλλά και το σχετικό μήνυμα που ενημερώνει το χρήστη ότι η φωτογραφία λήφθηκε.	46
Εικόνα 3. 4: Απεικονίζεται η οπτική απεικόνιση που πραγματοποιήθηκε από το Google Cloud’s Vision API στην εικόνα που λήφθηκε από το GUI. Φαίνεται ότι το Vision API αναγνώρισε ότι το περιεχόμενο της εικόνας είναι σχετικό με φαγητό.	48
Εικόνα 3. 5: Απεικονίζεται η οπτική απεικόνιση που πραγματοποιήθηκε από το Google Cloud’s Vision API στην εικόνα που λήφθηκε από το GUI. Παρουσιάζονται οι ετικέτες (labels) που επιστρέφονται από το Vision	

API σε συνδυασμό με την πιθανότητα που δίνει το API να αναγνώρισε σωστά την αντίστοιχη ετικέτα μέσα από την εικόνα. Φαίνονται μόνο τα πρώτα labels δηλαδή εκείνα με την μεγαλύτερη βαθμολογία (score)....	48
Εικόνα 3. 6: Ενδεικτική εικόνα ρυζιού από το σύνολο δεδομένων που εισάγεται στο GAN ώστε να εκπαιδευτεί αν έχει χαρακτηριστεί η φωτογραφία που τραβήχτηκε από το ψυγείο μας με ετικέτα “Rice”....	51
Εικόνα 3. 7: Ενδεικτική εικόνα λαχανικών από το σύνολο δεδομένων που εισάγεται στο GAN ώστε να εκπαιδευτεί αν έχει χαρακτηριστεί η φωτογραφία που τραβήχτηκε από το ψυγείο μας με ετικέτα “Vegetables”.....	51
Εικόνα 3. 8: Ενδεικτική εικόνα ψωμιού από το σύνολο δεδομένων που εισάγεται στο GAN ώστε να εκπαιδευτεί αν έχει χαρακτηριστεί η φωτογραφία που τραβήχτηκε από το ψυγείο μας με ετικέτα “Bread”.	52
Εικόνα 3. 9: Ενδεικτική εικόνα μακαρονιών από το σύνολο δεδομένων που εισάγεται στο GAN ώστε να εκπαιδευτεί αν έχει χαρακτηριστεί η φωτογραφία που τραβήχτηκε από το ψυγείο μας με ετικέτα “Pasta” .	52
Εικόνα 3. 10: Ενδεικτική εικόνα πατατών από το σύνολο δεδομένων που εισάγεται στο GAN ώστε να εκπαιδευτεί αν έχει χαρακτηριστεί η φωτογραφία που τραβήχτηκε από το ψυγείο μας με ετικέτα “Potatoes”.....	52
Εικόνα 3. 11: Ενδεικτική εικόνα κρέατος από το σύνολο δεδομένων που εισάγεται στο GAN ώστε να εκπαιδευτεί αν έχει χαρακτηριστεί η φωτογραφία που τραβήχτηκε από το ψυγείο μας με ετικέτα “Meat”...	52
Εικόνα 3. 12: Απεικονίζεται ο generator DCGAN. Μια ομοιόμορφη κατανομή 100 z προβάλλεται σε μικρή χωρική έκταση συνελκτικής αναπαράστασης με πολλούς χάρτες χαρακτηριστικών. Έπειτα μια σειρά τεσσάρων κλασματικών συνελίξεων μετατρέπουν αυτήν την αναπαράσταση υψηλού επιπέδου σε εικόνα 64 x 64 pixel. Επίσης σημαντική παρατήρηση είναι ότι δε χρησιμοποιούνται πλήρως συνδεδεμένα (fully connected) ή pooling στρώματα. Η εικόνα πάρθηκε από [51].	56
Εικόνα 3. 13: Η παραγόμενη εικόνα του Image-Generator GAN όταν χρησιμοποιούνται οι ήδη υπάρχουσες τιμές για όλες τις παραμέτρους.	57
Εικόνα 3. 14: Η παραγόμενη εικόνα του Image-Generator GAN για num_epochs = 50. Το GAN εκπαιδεύτηκε μόνο με 200 εικόνες λαχανικών.....	57
Εικόνα 3. 15: Η παραγόμενη εικόνα του Image-Generator GAN για num_epochs = 50 και n = 1 δηλαδή να δημιουργηθεί μια μόνο εικόνα. Το GAN εκπαιδεύτηκε μόνο με 200 εικόνες λαχανικών.....	58
Εικόνα 3. 16: Η παραγόμενη εικόνα του Image-Generator GAN για num_epochs = 100 και n = 1 δηλαδή να δημιουργηθεί μια μόνο εικόνα. Το GAN εκπαιδεύτηκε μόνο με 200 εικόνες λαχανικών.....	58
Εικόνα 3. 17: Η παραγόμενη εικόνα του Image-Generator GAN για num_epochs = 200 και n = 1 δηλαδή να δημιουργηθεί μια μόνο εικόνα. Το GAN εκπαιδεύτηκε μόνο με 200 εικόνες λαχανικών.....	59
Εικόνα 3. 18: Η παραγόμενη εικόνα του Image-Generator GAN για num_epochs = 300 και n = 1 δηλαδή να δημιουργηθεί μια μόνο εικόνα. Το GAN εκπαιδεύτηκε μόνο με 200 εικόνες λαχανικών.....	59
Εικόνα 3. 19: Η παραγόμενη εικόνα 128 x 128 pixels του Image-Generator GAN για num_epochs = 200 και n = 1 δηλαδή να δημιουργηθεί μια μόνο εικόνα. Το GAN εκπαιδεύτηκε μόνο με 200 εικόνες λαχανικών....	60
Εικόνα 3. 20: Η παραγόμενη εικόνα 128 x 128 pixels του Image-Generator GAN για num_epochs = 200 και n = 1. Το GAN εκπαιδεύτηκε μόνο με 200 εικόνες μακαρονιών.	61
Εικόνα 3. 21: Η παραγόμενη εικόνα 128 x 128 pixels του Image-Generator GAN για num_epochs = 200 και n = 1. Το GAN εκπαιδεύτηκε μόνο με 200 εικόνες ρυζιού.	61

Εικόνα 3. 22: Η παραγόμενη εικόνα 128 x 128 pixels του Image-Generator GAN για num_epochs = 200 και n = 1. Το GAN εκπαιδεύτηκε μόνο με 200 εικόνες ψωμιού.....	62
Εικόνα 3. 23: Η παραγόμενη εικόνα 128 x 128 pixels του Image-Generator GAN για num_epochs = 200 και n = 1. Το GAN εκπαιδεύτηκε μόνο με 200 εικόνες κρέατος.....	62
Εικόνα 3. 24: Η παραγόμενη εικόνα 128 x 128 pixels του Image-Generator GAN για num_epochs = 200 και n = 1. Το GAN εκπαιδεύτηκε μόνο με 200 εικόνες πατατών.....	63
Εικόνα 3. 25: Η παραγόμενη εικόνα 128 x 128 pixels του Image-Generator GAN για num_epochs = 200 και n = 1. Το GAN εκπαιδεύτηκε με 200 εικόνες μακαρονιών και 200 εικόνες πατατών.....	63
Εικόνα 3. 26: Η παραγόμενη εικόνα 128 x 128 pixels του Image-Generator GAN για num_epochs = 200 και n = 1. Το GAN εκπαιδεύτηκε με 200 εικόνες κρέατος, 200 εικόνες ρυζιού και 200 εικόνες λαχανικών.....	64
Εικόνα 3. 27: Η παραγόμενη εικόνα 128 x 128 pixels του Image-Generator GAN για num_epochs = 200 και n = 1. Το GAN εκπαιδεύτηκε με 200 εικόνες κρέατος, 200 εικόνες ρυζιού, 200 εικόνες ψωμιού, 200 εικόνες πατατών και 200 εικόνες λαχανικών.....	64
Εικόνα 4. 1: Διάγραμμα α) Στο αριστερά διάγραμμα απεικονίζονται οι τιμές για την ακρίβεια του μοντέλου (accuracy) από την πρώτη εποχή (1η) έως την τετρακοσιοστή (400ή) που ήταν και η τελευταία με learning rate = 0.0001 και batch = 32. β) Στο δεξιά διάγραμμα απεικονίζονται οι τιμές για τις απώλειες του μοντέλου (loss) για τις ίδιες εποχές με το διάγραμμα στα αριστερά. Η εκπαίδευση έγινε στις κλάσεις ψωμιού και μακαρονιών.....	67
Εικόνα 4. 2: Διάγραμμα α) Στο αριστερά διάγραμμα απεικονίζονται οι τιμές για την ακρίβεια του μοντέλου (accuracy) από την πρώτη εποχή (1η) έως την τετρακοσιοστή (400ή) που ήταν και η τελευταία με learning rate = 0.0001 και batch = 16. β) Στο δεξιά διάγραμμα απεικονίζονται οι τιμές για τις απώλειες του μοντέλου (loss) για τις ίδιες εποχές με το διάγραμμα στα αριστερά. Η εκπαίδευση έγινε στις κλάσεις ψωμιού και μακαρονιών.....	67
Εικόνα 4. 3: Διάγραμμα α) Στο αριστερά διάγραμμα απεικονίζονται οι τιμές για την ακρίβεια του μοντέλου (accuracy) από την πρώτη εποχή (1η) έως την τετρακοσιοστή (400ή) που ήταν και η τελευταία με learning rate = 0.001 και batch = 32. β) Στο δεξιά διάγραμμα απεικονίζονται οι τιμές για τις απώλειες του μοντέλου (loss) για τις ίδιες εποχές με το διάγραμμα στα αριστερά. Η εκπαίδευση έγινε στις κλάσεις ψωμιού και μακαρονιών.....	68
Εικόνα 4. 4: Διάγραμμα α) Στο αριστερά διάγραμμα απεικονίζονται οι τιμές για την ακρίβεια του μοντέλου (accuracy) από την πρώτη εποχή (1η) έως την πεντηκοστή (50ή) που ήταν και η τελευταία με learning rate = 0.001 και batch = 32. β) Στο δεξιά διάγραμμα απεικονίζονται οι τιμές για τις απώλειες του μοντέλου (loss) για τις ίδιες εποχές με το διάγραμμα στα αριστερά. Η εκπαίδευση έγινε στις κλάσεις ψωμιού και λαχανικών.....	75
Εικόνα 4. 5: Διάγραμμα α) Στο αριστερά διάγραμμα απεικονίζονται οι τιμές για την ακρίβεια του μοντέλου (accuracy) από την πρώτη εποχή (1η) έως την πεντηκοστή (50ή) που ήταν και η τελευταία με learning rate = 0.0001 και batch = 32. β) Στο δεξιά διάγραμμα απεικονίζονται οι τιμές για τις απώλειες του μοντέλου (loss) για τις ίδιες εποχές με το διάγραμμα στα αριστερά. Η εκπαίδευση έγινε στις κλάσεις ψωμιού και λαχανικών.....	75
Εικόνα 4. 6: Διάγραμμα α) Στο αριστερά διάγραμμα απεικονίζονται οι τιμές για την ακρίβεια του μοντέλου (accuracy) από την πρώτη εποχή (1η) έως την πεντηκοστή (50ή) που ήταν και η τελευταία με learning rate = 0.0001 και batch = 16. β) Στο δεξιά διάγραμμα απεικονίζονται οι τιμές για τις απώλειες του μοντέλου (loss) για τις ίδιες εποχές με το διάγραμμα στα αριστερά. Η εκπαίδευση έγινε στις κλάσεις ψωμιού και λαχανικών.....	76

Κατάλογος πινάκων

Πίνακας 3. 1: Παρουσιάζονται οι παράμετροι εκπαίδευσης του Image-Generator, οι default τιμές αυτών και η περιγραφή τους [52].	54
Πίνακας 3. 2: Παρουσιάζονται οι παράμετροι του Image-Generator για την δημιουργία παραγόμενων εικόνων, οι default τιμές αυτών και η περιγραφή τους [52].	55
Πίνακας 4. 1: Παρουσιάζονται οι τιμές για το True Positive, False Negative, False Positive και το True Negative για τις εποχές 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 210 και 220. Ως Positive ορίζονται οι εικόνες ψωμιού και ως Negative ορίζονται οι εικόνες μακαρονιών. Το μοντέλο κλήθηκε να κατηγοριοποιήσει 40 πραγματικές εικόνες ψωμιού και 40 πραγματικές εικόνες μακαρονιών (validation dataset).	69
Πίνακας 4. 2: Παρουσιάζονται οι τιμές για το True positive rate, False negative rate, False positive rate και το True negative rate για τις εποχές 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 210 και 220. Ως Positive ορίζονται οι εικόνες ψωμιού και ως Negative ορίζονται οι εικόνες μακαρονιών. Το μοντέλο κλήθηκε να κατηγοριοποιήσει 40 πραγματικές εικόνες ψωμιού και 40 πραγματικές εικόνες μακαρονιών (validation dataset). Έγινε στρωγγυλοποίηση δύο δεκαδικών ψηφίων.	70
Πίνακας 4. 3: Παρουσιάζονται οι τιμές για το Accuracy, το Balanced accuracy, το Positive Predictive value, Negative predict value και το F1 score για τις εποχές 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 210 και 220. Ως Positive ορίζονται οι εικόνες ψωμιού και ως Negative ορίζονται οι εικόνες μακαρονιών. Το μοντέλο κλήθηκε να κατηγοριοποιήσει 40 πραγματικές εικόνες ψωμιού και 40 πραγματικές εικόνες μακαρονιών (validation dataset). Έγινε στρωγγυλοποίηση δύο δεκαδικών ψηφίων.	71
Πίνακας 4. 4: Παρουσιάζονται οι τιμές για το True Positive, False Negative, False Positive και το True Negative για την εποχή 150. Ως Positive ορίζονται οι εικόνες ψωμιού και ως Negative ορίζονται οι εικόνες μακαρονιών. Το μοντέλο κλήθηκε να κατηγοριοποιήσει 50 παραγόμενες εικόνες ψωμιού και 50 παραγόμενες εικόνες μακαρονιών (test dataset).	72
Πίνακας 4. 5: Παρουσιάζονται οι τιμές για το True positive rate, False negative rate, False positive rate και το True negative rate για την εποχή 150. Ως Positive ορίζονται οι εικόνες ψωμιού και ως Negative ορίζονται οι εικόνες μακαρονιών. Το μοντέλο κλήθηκε να κατηγοριοποιήσει 50 παραγόμενες εικόνες ψωμιού και 50 παραγόμενες εικόνες μακαρονιών (test dataset). Έγινε στρωγγυλοποίηση δύο δεκαδικών ψηφίων.	73
Πίνακας 4. 6: Παρουσιάζονται οι τιμές για το Accuracy, το Balanced accuracy, το Positive Predictive value, Negative predict value και το F1 score για την εποχή 150. Ως Positive ορίζονται οι εικόνες ψωμιού και ως Negative ορίζονται οι εικόνες μακαρονιών. Το μοντέλο κλήθηκε να κατηγοριοποιήσει 50 παραγόμενες εικόνες ψωμιού και 50 παραγόμενες εικόνες μακαρονιών (test dataset). Έγινε στρωγγυλοποίηση δύο δεκαδικών ψηφίων.	73
Πίνακας 4. 7: Παρουσιάζονται οι τιμές για το True Positive, False Negative, False Positive και το True Negative για τις εποχές 10, 15, 20, 25, 30, 35, 40, 45 και 50. Ως Positive ορίζονται οι εικόνες ψωμιού και ως Negative ορίζονται οι εικόνες λαχανικών. Το μοντέλο κλήθηκε να κατηγοριοποιήσει 40 πραγματικές εικόνες ψωμιού και 40 πραγματικές εικόνες λαχανικών (validation dataset).	77
Πίνακας 4. 8: Παρουσιάζονται οι τιμές για το True positive rate, False negative rate, False positive rate και το True negative rate για τις εποχές 10, 15, 20, 25, 30, 35, 40, 45 και 50. Ως Positive ορίζονται οι εικόνες ψωμιού και ως Negative ορίζονται οι εικόνες λαχανικών. Το μοντέλο κλήθηκε να κατηγοριοποιήσει 40 πραγματικές εικόνες ψωμιού και 40 πραγματικές εικόνες λαχανικών (validation dataset). Έγινε στρωγγυλοποίηση δύο δεκαδικών ψηφίων.	78

Πίνακας 4. 9: Παρουσιάζονται οι τιμές για το Accuracy, το Balanced accuracy, το Positive Predictive value, Negative predict value και το F1 score για τις εποχές 10, 15, 20, 25, 30, 35, 40, 45 και 50. Ως Positive ορίζονται οι εικόνες ψωμιού και ως Negative ορίζονται οι εικόνες λαχανικών. Το μοντέλο κλήθηκε να κατηγοριοποιήσει 40 πραγματικές εικόνες ψωμιού και 40 πραγματικές εικόνες λαχανικών (validation dataset). Έγινε στρογγυλοποίηση δύο δεκαδικών ψηφίων. 79

Πίνακας 4. 10: Παρουσιάζονται οι τιμές για το True Positive, False Negative, False Positive και το True Negative για την εποχή 20. Ως Positive ορίζονται οι εικόνες ψωμιού και ως Negative ορίζονται οι εικόνες λαχανικών. Το μοντέλο κλήθηκε να κατηγοριοποιήσει 50 παραγόμενες εικόνες ψωμιού και 50 παραγόμενες εικόνες λαχανικών (test dataset). 80

Πίνακας 4. 11: Παρουσιάζονται οι τιμές για το True positive rate, False negative rate, False positive rate και το True negative rate για την εποχή 20. Ως Positive ορίζονται οι εικόνες ψωμιού και ως Negative ορίζονται οι εικόνες λαχανικών. Το μοντέλο κλήθηκε να κατηγοριοποιήσει 50 παραγόμενες εικόνες ψωμιού και 50 παραγόμενες εικόνες λαχανικών (test dataset). Έγινε στρογγυλοποίηση δύο δεκαδικών ψηφίων. 80

Πίνακας 4. 12: Παρουσιάζονται οι τιμές για το Accuracy, το Balanced accuracy, το Positive Predictive value, Negative predict value και το F1 score για την εποχή 20. Ως Positive ορίζονται οι εικόνες ψωμιού και ως Negative ορίζονται οι εικόνες λαχανικών. Το μοντέλο κλήθηκε να κατηγοριοποιήσει 50 παραγόμενες εικόνες ψωμιού και 50 παραγόμενες εικόνες λαχανικών (test dataset). Έγινε στρογγυλοποίηση δύο δεκαδικών ψηφίων. 81

Κεφάλαιο 1: Εισαγωγή

1.1 Ιστορική αναδρομή

Από τα αρχαία χρόνια ο άνθρωπος προσπαθεί να εξελίσσεται και να βελτιώνει με κάθε τρόπο την ποιότητα ζωής αλλά και την καθημερινότητά του. Αυτός ήταν και ο βασικός λόγος που αναπτύχθηκε και συνεχίζει να αναπτύσσεται σε τέτοιο βαθμό η τεχνολογία. Οι υπολογιστές κατέχουν πρωταγωνιστικό ρόλο στην σύγχρονη κοινωνία και εισχωρούν μέρα με την μέρα όλο πιο βαθιά στην ζωή μας [1]. Μάλιστα σύμφωνα με το αγγλικό λεξικό του Oxford (Oxford English Dictionary) η πρώτη γνωστή χρήση της λέξης υπολογιστής (computer) έγινε το 1613 σε ένα βιβλίο γραμμένο από τον Richard Braithwait. Ο Άγγλος συγγραφέας ουσιαστικά αναφερόταν στον ανθρώπινο υπολογιστή (human computer) και ειδικότερα σε έναν άνθρωπο ο οποίος θα έκανε υπολογισμούς [2]. Ωστόσο ο πρώτος κλασικός σύγχρονος υπολογιστής φαίνεται να είναι ο IBM 3070 ο οποίος τοποθετήθηκε στην NASA από πολλούς έμπειρους ανθρώπους πάνω στον κλάδο των υπολογιστών ώστε να καταφέρει να συναρμολογηθεί και να δουλέψει σωστά. Υποσχόταν να κάνει σε πολύ πιο γρήγορο χρόνο υπολογισμούς που έως τώρα συνήθιζαν να κάνουν οι άνθρωποι [2]. Μέρα με την μέρα η τεχνολογία εξελίσσεται, οι επεξεργαστές των υπολογιστών γίνονται όλο και καλύτεροι και η ανάγκη για να αναπτυχθεί ακόμα περισσότερο η τεχνολογία πολλαπλασιάζεται. Έχει εισβάλει σε πολλούς τομείς της καθημερινότητας μας η τεχνολογία και γενικότερα οι υπολογιστές έχουν γίνει απαραίτητοι.

Από τα αυτοκίνητα που υιοθετούν όλο και περισσότερα αυτοματοποιημένα συστήματα μέχρι την ιατρική που χρησιμοποιεί ρομπότ και μηχανές για την διεκπεραίωση κάποιας χειρουργικής επέμβασης, φαίνεται η σημαντική συμβολή της τεχνολογίας. Παρατηρείται επομένως η ενσωμάτωση υπολογιστών σε διάφορα αντικείμενα. Θέλοντας λοιπόν να βελτιωθεί και η καθημερινότητα του ανθρώπου αλλά και η ποιότητα ζωής του [3], δημιουργήθηκαν οι “έξυπνες συσκευές”. Οι “έξυπνες συσκευές” ανήκουν στην κατηγορία του Διαδικτύου των Πραγμάτων (Internet of Things). Οι υπολογιστές και η τεχνολογία δηλαδή ενσωματώθηκε σε απλά καθημερινά αντικείμενα με σκοπό να δημιουργήσει ένα ευκολότερο περιβάλλον στον άνθρωπο [4]. Χάρη στην ταχύτατη ανάπτυξη της τεχνολογίας και την εύκολη χρήση του διαδικτύου οι “έξυπνες συσκευές” φαίνεται να έχουν εισέλθει στα σπίτια των ανθρώπων, με απώτερο σκοπό την δημιουργία του “έξυπνου σπιτιού”. Όπως εύκολα αντιλαμβάνεται κανείς το λεγόμενο “έξυπνο σπίτι” είναι ένα πολλά υποσχόμενο επίτευγμα της τεχνολογίας των υπολογιστών, το οποίο θα αποτελείται εξ’ ολοκλήρου από “έξυπνες συσκευές” που θα μπορούν να επικοινωνούν μεταξύ τους αλλά και να ελέγχονται από τον ιδιοκτήτη με την χρήση του διαδικτύου και με την βοήθεια αισθητήρων [5]. Κάποια παραδείγματα “έξυπνων συσκευών” που υπάρχουν ήδη στα περισσότερα σπίτια είναι οι “έξυπνες τηλεοράσεις”, τα κινητά τηλέφωνα, τα “έξυπνα ρολόγια”, οι “έξυπνοι εκτυπωτές” (3d printer) και άλλα [6].

1.2 Σκοπός της παρούσας εργασίας

Η κουζίνα είναι ένας βασικός χώρος μέσα στο σπίτι που η τεχνολογία θα ευδοκιμήσει λόγω του τρόπου ζωής των ανθρώπων σήμερα. Η τεχνολογία τείνει να δημιουργήσει μια κουζίνα ή οποία θα αποτελείται εξολοκλήρου από “έξυπνες συσκευές”, οι οποίες θα ελέγχονται από τον ιδιοκτήτη τους και θα συνεργάζονται μεταξύ τους για την ομαλή λειτουργία τους. Με λίγα λόγια όλες οι “έξυπνες συσκευές”, γενικότερα του “έξυπνου σπιτιού” και όχι μόνο της “έξυπνης κουζίνας”, θα έχουν την δυνατότητα να είναι εν μέρη αυτόνομες καθώς θα εκτελούν τις λειτουργίες, για τις οποίες έχουν σχεδιαστεί, από μόνες τους. Επίσης θα δίνουν στον χρήστη επιπλέον δυνατότητες και λειτουργίες κάνοντας του την καθημερινότητα ευκολότερη. Πάνω σε αυτή την βάση έγινε αντιληπτή η χρησιμότητα ενός “έξυπνου ψυγείου”. Ο περιορισμένος ελεύθερος χρόνος και οι αυξημένες προσδοκίες για υγιεινές συνήθειες στις διατροφικές επιλογές μας, επιβάλει την δημιουργία του “έξυπνου ψυγείου” στον κόσμο της τεχνολογίας. Σε αυτήν λοιπόν την εργασία παρουσιάζεται η ιδέα αυτή. Ένα “έξυπνο ψυγείο” που θα έχει την δυνατότητα να αναγνωρίζει τι τρόφιμα και υλικά περιέχει, εμφανίζοντας εικόνες από πιθανή σύνθεση μερικών ή όλων των υλικών που υπάρχουν ήδη στο ψυγείο. Με λίγα λόγια μας προτείνει τι μπορεί να μαγειρέψει ο ενδιαφερόμενος με τα τρόφιμα που διαθέτει. Μέσω υπολογιστικών μεθόδων που θα είναι ενσωματωμένα, υποθετικά, στο ψυγείο, ο χρήστης θα μπορεί να πάρει μια ιδέα για το τί να μαγειρέψει. Θα παίρνει λοιπόν σαν αποτέλεσμα τη σύνθεση νέων εικόνων που θα απεικονίζουν το αποτέλεσμα της “μαγειρικής” των υλικών που διαθέτει το ψυγείο του. Ουσιαστικά

επιθυμούμε να βοηθήσουμε το χρήστη, και να έχει οπτική επαφή με το γεύμα του, δηλαδή το τι θα μπορούσε να δημιουργήσει με τα υλικά που διαθέτει εκείνη την στιγμή. Επίσης επιθυμούμε οι διατροφικές συστάσεις από το “έξυπνο ψυγείο” να οδηγούν σε μια υγιεινή και ισορροπημένη διατροφή [7] [8].

Συγκεκριμένα η λειτουργία του προτεινόμενου “έξυπνου ψυγείου” θα αποτελείται από τα ακόλουθα στάδια: πρώτα, θα φωτογραφίζεται το εσωτερικό του ψυγείου με μια φορητή κάμερα. Έπειτα μέσω της μηχανικής όρασης (computer vision) αναλύεται η φωτογραφία που τραβήχτηκε και γνωστοποιούνται τα αντικείμενα, δηλαδή τα υλικά, που υπάρχουν στο ψυγείο σε γραπτή μορφή. Έχοντας συλλέξει και ξεχωρίσει τα υλικά της εικόνας, τα οποία είναι εφικτό να χρησιμοποιηθούν για μαγείρεμα, τα χρησιμοποιούμε ως είσοδο σε ένα νευρωνικό δίκτυο. Συγκεκριμένα το νευρωνικό δίκτυο αυτό λέγεται Generative Adversarial Networks (GANs). Αφού λοιπόν εκπαιδευτεί το GAN με τα κατάλληλα σύνολα δεδομένων των τροφών που διαθέτει ο χρήστης, δημιουργεί την νέα εικόνα από την σύνθεση των συστατικών που είναι διαθέσιμα. Έτσι, καταφέρνει να μετατρέψει απλές εικόνες συστατικών φαγητού, σε εικόνες που περιλαμβάνουν πολλά υλικά φαγητού. Με λίγα λόγια μια εντελώς καινούργια εικόνα γεννιέται και σε αυτή αποτυπώνεται η σύνθεση μερικών ή όλων των διαθέσιμων υλικών που θα υπήρχαν, εάν είχαν μαγειρευτεί. Έπειτα αξιολογείται η αποτελεσματικότητα του αλλά και γενικότερα η νεοσύστατη εικόνα που δημιουργήθηκε.

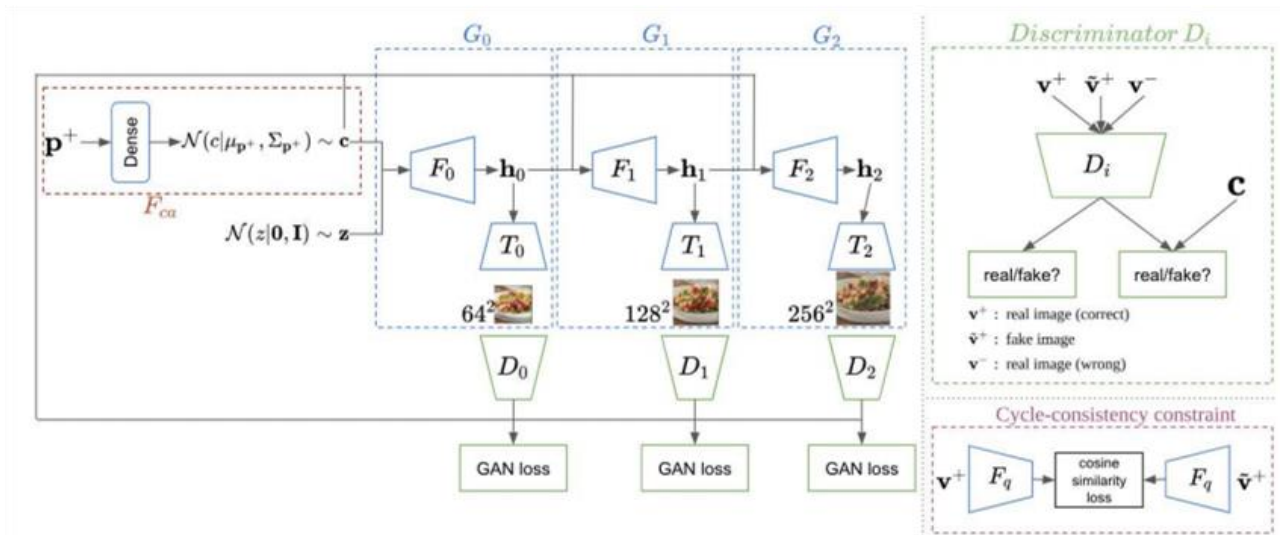
Η λειτουργία αυτή συμβάλει στην εξοικονόμηση χρημάτων, χρόνου αλλά και στην μείωση των τροφών που καταλήγουν στα απορρίμματα. Καταφέρνοντας να μαγειρέψει ο χρήστης με όσα διαθέτει ήδη, αποφεύγει την εύκολη λύση ενός έτοιμου αγορασμένου γεύματος που κατά πάσα πιθανότητα θα ήταν και πιο ανθυγιεινό. Επιπλέον χρησιμοποιεί στο έπακρο τα ήδη αγορασμένα υλικά ενώ έχει οπτική επαφή με το αποτέλεσμα πριν καν το δημιουργήσει [6].

1.3 Σχετική βιβλιογραφία

Η ιδέα του “έξυπνου ψυγείου” δεν είναι καινούργια αφού στο παρελθόν έχουν υπάρξει παρόμοιες εργασίες που παρουσιάζουν αντίστοιχες προσεγγίσεις ενός “έξυπνου ψυγείου”. Παρακάτω παρουσιάζονται μερικές από τις προσεγγίσεις διάφορων εργασιών που έχουν πραγματοποιηθεί.

1.3.1 Cook-GAN

Το Cook-GAN προτάθηκε από ερευνητές του πανεπιστημίου Rutgers (Rutgers University) και το τμήμα της Αγγλίας Samsung AI Center. Σκοπός του Cook-GAN είναι η συμβολή του στην δημιουργία του “έξυπνου ψυγείου”. Επιτρέπει στον χρήστη να εισάγει μία λίστα με υλικά που θέλει να απεικονιστούν στην νέα εικόνα που θα δημιουργηθεί από το Cook-GAN. Ουσιαστικά το αποτέλεσμα της εφαρμογής αυτής θα είναι μια καινούργια εικόνα ενός γεύματος φαγητού. Συγκεκριμένα η εργασία συνθέτει ρεαλιστικές εικόνες φαγητού, προερχόμενες από μία λίστα με συστατικά. Για την υλοποίηση χρησιμοποιούνται βαθιά μοντέλα παραγωγικής μάθησης (generative deep models). Η καινοτομία του Cook-GAN είναι η δημιουργία εικόνων που αφορά τα συστατικά του τελικού αποτελέσματος, που θα συνθέσουν ολόκληρο το γεύμα. Για παράδειγμα εάν σκοπός είναι να δημιουργηθεί τελικώς η εικόνα μιας πίτσας, πρώτα δημιουργούνται εικόνες που απεικονίζουν την μαγειρεμένη ντομάτα, κρεμμύδι, τυρί κ.λπ.. Στην συνέχεια αυτές οι εικόνες των συστατικών χρησιμοποιούνται για την δημιουργία και εκπαίδευση του νευρωνικού δικτύου που θα καταφέρει να συνθέσει ολόκληρο το γεύμα του φαγητού. Σε αντίθεση με προηγούμενες μελέτες που δημιούργησαν εικόνες χαμηλής ανάλυσης (π.χ. 128×128 pixel) ή μόνο εικόνες συγκεκριμένων τύπων τροφίμων (π.χ. πίτσα), το νέο μοντέλο μπορεί να δημιουργήσει εικόνες γεύματος πολλών τύπων και συστατικών τροφίμων. Έτσι επιτυγχάνεται μια πλήρης εικόνα που αντιστοιχεί σε ένα γεύμα και περιλαμβάνει τα απαραίτητα συστατικά. Οι ερευνητές αντλούν δεδομένα από ένα σύνολο δεδομένων που λέγεται “Recipe1M” [9] και περιλαμβάνει περίπου ένα εκατομμύριο συνταγές με τίτλους, οδηγίες, συστατικά και εικόνες. Δημιούργησαν ένα υποσύνολο 402.760 εικόνων που έχουν από ένα έως είκοσι συστατικά και αντίστοιχα από μία έως είκοσι οδηγίες για την συνταγή τους. Τα δεδομένα, δηλαδή οι εικόνες χωρίστηκαν σε επιμέρους υποσύνολα ώστε το 70 τοις εκατό αυτών να χρησιμοποιηθεί για την εκπαίδευση του νευρωνικού μοντέλου, το 15 τοις εκατό για επικύρωση (validation) των αποτελεσμάτων της εκπαίδευσης αυτού και 15 τοις εκατό ως σύνολο δοκιμών (testing), χρησιμοποιώντας το πολύ πέντε εικόνες από κάθε συνταγή [10].



Εικόνα 1. 1: Το μοντέλο που απεικονίζει την αξιολόγηση των εικόνων και αν αυτές είναι ρεαλιστικές ή όχι (discriminator) [10].

1.3.2 Gan-stronomy: Generative cooking with conditional DCGANs.

Το Gan-stronomy συμβάλει στην δημιουργική μαγειρική (generative cooking) και με την βοήθεια ενός παραγωγικού μοντέλου (generative model) που χρησιμοποιεί μαγειρικές συνταγές δημιουργεί μια νέα ρεαλιστική εικόνα ενός τελικού αποτελέσματος μαγειρικής, έχοντας ως είσοδο τα υλικά και τις οδηγίες παρασκευής του πιάτου. Παρόλο που δεν υπάρχει εκτενής βιβλιογραφία όσον αφορά την δημιουργική μαγειρική (generative cooking), οι δημιουργοί του Gan-stronomy βρήκαν και χρησιμοποίησαν το Im2Recipe και τα δίκτυα DCGAN [11]. Συγκεκριμένα οι ερευνητές του Im2Recipe δημιούργησαν το RecipeIM, μια συλλογή από παραπάνω από ένα εκατομμύριο συνταγές και οκτακόσιες χιλιάδες φωτογραφίες φαγητών. Αυτή η μεγαλύτερη δημόσια διαθέσιμη συλλογή παρέχει την δυνατότητα εκπαίδευσης σε υψηλής χωρητικότητας μοντέλα με πολυμορφικά δεδομένα (multi-modal data). Έτσι εκπαίδευσαν ένα νευρωνικό δίκτυο ώστε να ανακτά υπάρχουσες εικόνες και συνταγές φαγητών. Ειδικότερα χρησιμοποιήθηκαν τα DCGAN (deep convolutional generative adversarial networks) δίκτυα, που μπορούν να συνθέσουν νέες ρεαλιστικές εικόνες. Επιλέχθηκαν συνταγές που είχαν τουλάχιστον μία φωτογραφία που τους αντιστοιχούσε από την συλλογή RecipeIM. Επιπλέον έγινε ένας διαχωρισμός των συνταγών ώστε κάθε συνταγή να αντιστοιχεί σε μια συγκεκριμένη κατηγορία σύμφωνα με το περιεχόμενο της κάθε μιας. Κάποιες κατηγορίες είναι οι “σαλάτα κοτόπουλου” και “κέικ σοκολάτας”. Επισημαίνεται ότι το μοντέλο εκπαιδεύτηκε σε 49.800 παραδείγματα για 90 εποχές (epochs) με μέγεθος παρτίδας 64 (batch size) και $\alpha\text{MSE} / \alpha\text{BCE} = 0$. Παρατηρήθηκε επίσης ότι η αύξηση του αMSE εμφανίζει πιο θολές τις εικόνες, ενώ καλύτερα αποτελέσματα εντοπίζονται όταν ο δείκτης αυτός έχει την τιμή μηδέν. Βέβαια εξετάζοντας την ποιότητα των εικόνων συγκρίνοντάς τις παραγόμενες με τις αληθινές, παρατηρήθηκε προφανώς μεγάλη διαφορά. Ωστόσο το σχήμα, οι χρωματισμοί και τα γενικότερα χαρακτηριστικά είναι σε πολύ καλό επίπεδο [11].

{boneless chicken breast halves, vinaigrette, vegetable oil, mixed salad greens, grape tomatoes, ...}



{large shrimp, kosher salt, mint, red wine, vinegar, whole wheat pita bread, Greek yogurt, ...}



Εικόνα 1. 2: Αποτελέσματα του Gan-stronomy [11].

1.3.3 A Food Dish Image Generation Framework Based on Progressive Growing GANs

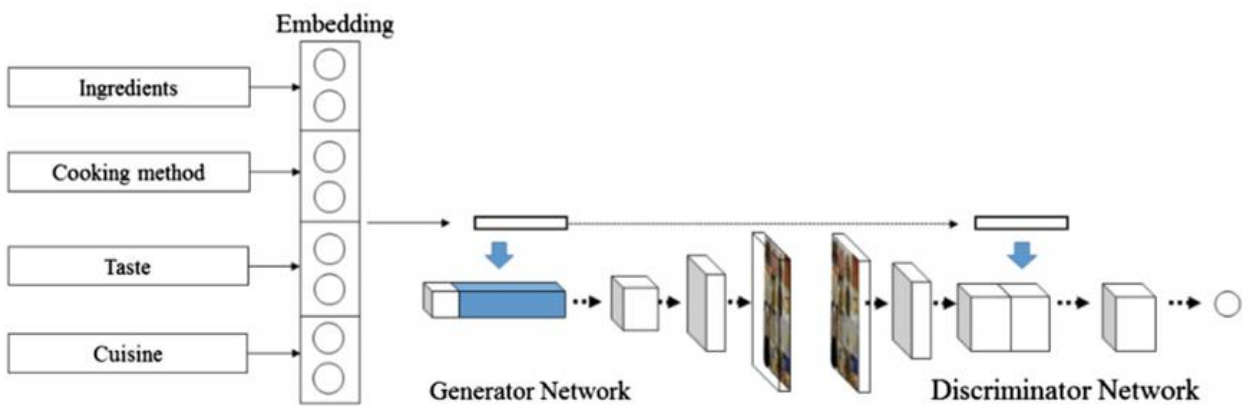
Η μέθοδος μαγειρέματος, τα συστατικά που περιλαμβάνει, η γεύση αλλά και η κουζίνα που ανήκει είναι χαρακτηριστικά που λαμβάνονται υπόψη στη συγκεκριμένη εργασία ώστε η παραγόμενη εικόνα που θα δημιουργηθεί να είναι φανερά βελτιωμένη. Επιπλέον όλα αυτά τα επιπλέον χαρακτηριστικά, λαμβάνονται υπόψη σε όλη την διάρκεια εκπαίδευσης. Ως σύνολο δεδομένων οι δημιουργοί χρησιμοποίησαν το Food-121 το οποίο περιλαμβάνει ονόματα τροφίμων, συστατικών, τρόπους μαγειρέματος, γεύσεις και το είδος της κουζίνας. Συγκεκριμένα περιέχει 121.478 κομμάτια πληροφοριών και εικόνων φαγητού. Τα αποτελέσματα άφησαν ικανοποιημένους τους δημιουργούς αφού το χρησιμοποιούμενο παραγωγικό δίκτυο μπορεί να παράγει εικόνες τροφίμων με πολύπλοκες λεπτομέρειες. Επιπροσθέτως απέκτησε υψηλή απόδοση συγκριτικά με άλλα μοντέλα πάνω στο συγκεκριμένο σύνολο δεδομένων (Food-121). Επίσης φαίνεται ότι το μοντέλο δεν είναι τόσο αποτελεσματικό στην δημιουργία του φόντου του φαγητού ίσως γιατί δεν υπάρχει αυτή η παράμετρος στο σύνολο δεδομένων Food-121 που εκπαιδεύεται [12].

Στο μοντέλο αυτό χρησιμοποιούνται τα “υπό όρους” GANs (conditional GANs) που είναι μια επέκταση των GANs. Η διαφορά μεταξύ των GANs και CGANs είναι ότι τα CGANs προσθέτουν ένα “υπό όρους” διάνυσμα εισόδου y στον τυχαίο θόρυβο z . Με αυτή την μέθοδο οι εικόνες φαίνονται πιο ρεαλιστικές και έχουν καλύτερη ανάλυση. Συγκεκριμένα η αρχιτεκτονική του μοντέλου αυτού περιλαμβάνει τρία μέρη όπως φαίνεται και στην εικόνα παρακάτω. Τα μέρη αυτά είναι ο κωδικοποιητής του κειμένου του φαγητού (food text encoder), το δίκτυο του generator και το δίκτυο του discriminator. Ο κωδικοποιητής του κειμένου του φαγητού αρχικά μετατρέπει τις πληροφορίες (τα συστατικά, την μέθοδο μαγειρέματος, τις πληροφορίες για την γεύση και το είδος της κουζίνας) σε διανύσματα και έπειτα ενώνει αυτά τα διανύσματα και τα τροφοδοτεί στο νευρωνικό δίκτυο του generator. Έπειτα όπως είναι αναμενόμενο ο generator δημιουργεί την συνθετική εικόνα η οποία αξιολογείται από το νευρωνικό δίκτυο του discriminator. Δηλαδή την συγκρίνει με μια πραγματική και προβλέπει την πιθανότητα να είναι πλαστή [12].

Ειδικότερα ο κωδικοποιητής του κειμένου των τροφίμων μαθαίνει μια παράσταση word2vec σε επίπεδο συστατικών ώστε να διαχειριστεί τις πληροφορίες που αφορούν τα συστατικά. Το word2vec είναι μια από τις πιο δημοφιλείς τεχνικές για την εκμάθηση ενσωματωμένων λέξεων χρησιμοποιώντας ένα νευρωνικό δίκτυο (shallow neural network). Ουσιαστικά ο αλγόριθμος αυτός προσπαθεί να προβλέψει το γενικότερο πλαίσιο της λέξης-στόχου [13]. Άρα εδώ μέσω του word2vec ο κωδικοποιητής παίρνει τα συστατικά και τα αποκωδικοποιεί ώστε να μπορέσει να τα χρησιμοποιήσει. Η μέθοδο μαγειρέματος κωδικοποιείται από ένα “skip-instruction” μοντέλο. Ειδικότερα το skip-instruction είναι ένα είδος μοντέλου που προβλέπει τις επόμενες οδηγίες και χρησιμοποιείται για να γίνει η αναπαράσταση των οδηγιών μαγειρέματος. Επίσης βασίζεται στην τεχνική “skip-thoughts” τεχνική. Δηλαδή έναν “unsupervised” κωδικοποιητή προτάσεων εκμάθησης. Ουσιαστικά η τεχνική αυτή κωδικοποιεί μια πρόταση και χρησιμοποιεί αυτή την κωδικοποίηση ως οδηγό για την αποκωδικοποίηση των προηγούμενων και των επόμενων προτάσεων. Με λίγα λόγια το “skip-thought” αντιστοιχίζει προτάσεις παρόμοια σημασιολογικά σε παρόμοιες διανυσματικές παραστάσεις. Η skip-instruction μέθοδος προσθέτει μια αρχή και ένα τέλος στις οδηγίες και χρησιμοποιεί ένα LSTM αντί για μια επαναλαμβανόμενη πύλη (GRU-gated recurrent unit) [12]. Το LSTM-long short-term memory ή

αλλιώς μακροχρόνια βραχυπρόθεσμη μνήμη είναι μία τεχνητή επαναλαμβανόμενη αρχιτεκτονική νευρωνικού δικτύου (RNN) που χρησιμοποιείται στον τομέα βαθιάς μάθησης (deep learning). Σε αντίθεση με τα τυπικά νευρωνικά δίκτυα feedforward το LSTM έχει συνδέσεις ανατροφοδότησης. Εκτός από μεμονωμένα σημεία δεδομένων (όπως εικόνες για παράδειγμα) μπορεί να επεξεργαστεί και ολόκληρες ακολουθίες δεδομένων (όπως βίντεο) [14]. Από την άλλη οι GRUs είναι ένας μηχανισμός πύλης σε επαναλαμβανόμενα νευρωνικά δίκτυα. Το GRU μοιάζει με LSTM με την διαφορά ότι έχει λιγότερες παραμέτρους και δεν έχει πύλη εξόδου [15]. Για την γεύση και για το είδος της κουζίνας χρησιμοποιείται η αναπαράσταση word2vec. Έπειτα όλες αυτές οι πληροφορίες συνδυάζονται για την πλήρη εικόνα της κατάστασης του επιθυμητού φαγητού [12].

Το μοντέλο εκπαιδεύεται με προοδευτικά αυξανόμενο τρόπο όπως λειτουργούν γενικά τα GANs. Ξεκινά την εκπαίδευση του δημιουργού και του διαχωριστή με εικόνες ανάλυσης 4 επί 4 pixels και σταδιακά προστίθενται επιπλέον στρώματα και αξιολογούνται. Ειδικότερα ο λανθάνων φορέας (latent vector) μαζί με τις πληροφορίες του φαγητού που αναλύθηκαν προηγουμένως μπαίνει σαν είσοδος στον δημιουργό και δημιουργεί μια εικόνα χαμηλής ποιότητας. Έπειτα η παραγόμενη εικόνα, οι πληροφορίες του φαγητού και η πραγματική εισάγονται στον διαχωριστή και παράγεται η πιθανότητα να είναι αληθινή η παραγόμενη εικόνα. Στις επόμενες εποχές (epoch) το GAN δημιουργεί εικόνες υψηλότερης ανάλυσης και έτσι η ποιότητα αυτών βελτιώνεται αλλά και απεικονίζει περισσότερες λεπτομέρειες [12].



Εικόνα 1. 3: Αναπαράσταση της αρχιτεκτονικής [12].

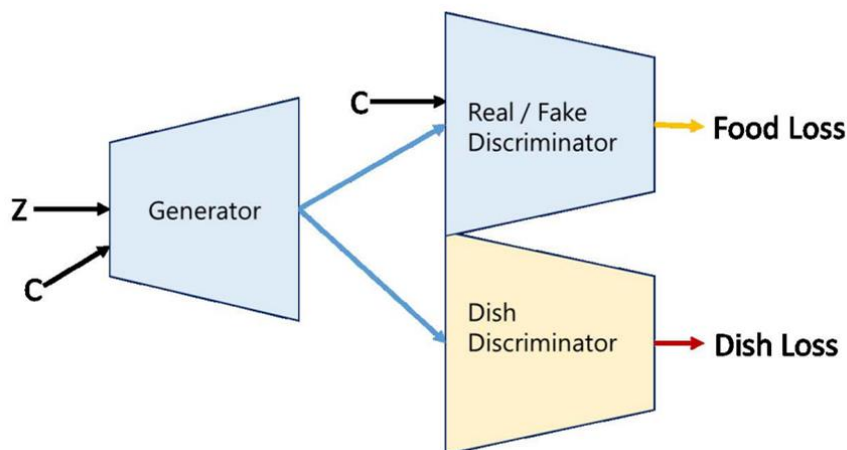
1.3.4 Food Image Generation using A Large Amount of Food Images with Conditional GAN: RamenGAN and RecipeGAN

Σε αυτή την εργασία παρουσιάστηκε η δημιουργία εικόνων χρησιμοποιώντας τα “υπό όρους” (conditional) Generative Adversarial Networks (CGANs) έχοντας μεγάλης κλίμακας σύνολα δεδομένων. Η εργασία αυτή αποτελείται από δύο παραγωγικά ανταγωνιστικά δίκτυα τα οποία παρουσιάζονται παρακάτω και κάθε ένα δίκτυο χρησιμοποιεί διαφορετικό σύνολο δεδομένων.

1.3.4.1 RamenGAN

Για το GAN αυτό χρησιμοποιήθηκε η πρώτη συλλογή δεδομένων, η “Ramen image dataset”, όπου οι συγγραφείς της εργασίας συνέλεξαν εικόνες από το twitter και αφαίρεσαν τον θόρυβο. Συγκεκριμένα το σύνολο δεδομένων αυτό αποτελείται από 15.900 εικόνες οι οποίες χωρίστηκαν σε έξι κατηγορίες που κάθε κατηγορία έχει συγκεκριμένη ετικέτα (label). Αυτές οι εικόνες σε συνδυασμό με την λεζάντα τους, εκπαίδευσαν το “υπό όρους” (conditional) GAN που για να εκπαιδευτεί απαιτεί κάθε εικόνα εκπαίδευσης να χαρακτηρίζεται από μία ετικέτα. Για την εκπαίδευση του “υπό όρους” GAN χρησιμοποιήθηκε ένα ενιαίο διάνυσμα για τα ramen. Ωστόσο το “υπό όρους” διάνυσμα (conditional vector) έχει έξι διαστάσεις αφού το

ramen χωρίζεται σε 6 κατηγορίες. Αφού εκπαιδευτεί δημιουργεί νέες εικόνες. Οι συντάκτες θέλοντας να βελτιώσουν τις εικόνες που δημιουργούνται κάνουντάς τες πιο καθαρές πρότειναν μια επέκταση του GAN. Σκοπός τους ήταν το φαγητό της φωτογραφίας να βρίσκεται μέσα σε ένα στρογγυλό/οβάλ πιάτο. Για να επιτευχθεί αυτό εισήγαγαν έναν επιπλέον discriminator που ελέγχει εάν το πιάτο είναι στρογγυλό. Αυτό το δίκτυο εκπαιδεύεται με σχήματα οβάλ ως αποδεκτά. Οπότε τελικώς η νέα εικόνα που δημιουργείται από τον generator περνάει από δύο discriminator, ο ένας ελέγχει την πιθανότητα να είναι αληθινή η εικόνα αυτή και ο δεύτερος ελέγχει κατά πόσο είναι σωστό το σχήμα του πιάτου. Από τα αποτελέσματα διαπιστώνεται ότι οι εικόνες οι οποίες έχουν παραχθεί από την εκπαίδευση που έγινε με την παρουσία του discriminator του πιάτου έχουν στρογγυλό σχήμα πιάτου και είναι παρόμοιες με τις πραγματικές. Αντίθετα εκείνες που προήλθαν από το δίκτυο που δεν είχε τον επιπλέον discriminator φαίνεται να έχουν παραμορφωμένο σχήμα πιάτου [16].



Εικόνα 1. 4: Απεικονίζονται τα τρία νευρωνικά δίκτυα, δηλαδή ο generator και οι δύο discriminators [16].

1.3.4.2 RecipeGAN

Για το δεύτερο GAN, το RecipeGAN χρησιμοποιήθηκε το σύνολο δεδομένων που λέγεται CookPad και περιλαμβάνει περισσότερες από ένα εκατομμύριο συνταγές με εικόνες φαγητών και πληροφορίες γύρω από τις συνταγές όπως συστατικά, μέθοδος μαγειρέματος και φυσικά τα ονόματα των συνταγών. Αυτό το σύνολο δεδομένων χρησιμοποιείται ως σύνολο εκπαίδευσης για την δημιουργία εικόνων από πληροφορίες συστατικών. Χρησιμοποιήθηκαν τα 127.690 είδη συνταγών που περιέχουν πληροφορίες συστατικών χωρίς όμως καρυκεύματα. Σε αυτό το μοντέλο η δημιουργία εικόνων πρέπει να παραχθούν από τα δεδομένα της συνταγής, δηλαδή τα συστατικά της, τα οποία έχουν μεγαλύτερη ποικιλία. Κατά την διάρκεια της εκπαίδευσης οι ‘‘υπό όρους’’ πληροφορίες (conditional information) που χρησιμοποιούνται για το δεύτερο σύνολο δεδομένων είναι ετικέτες που περιγράφουν συστατικά της συνταγής. Τέτοια εκπαίδευση δε μπορούσε να επιτευχθεί με τα απλά GANs γιατί παίρνουν ως είσοδο μόνο τα διανύσματα θορύβου. Η σημαντικότερη δυσκολία της εκπαίδευσης του κλασσικού GAN είναι ότι οι εικόνες που δημιουργούνται τείνουν να μοιάζουν μεταξύ τους και αυτό προκαλεί την κατάρρευση του μοντέλου αυτού. Για να αντιμετωπιστεί αυτό το πρόβλημα γίνεται χρήση του WGAN-GP που αντιμετωπίζει το πρόβλημα της κατάρρευσης ενός μοντέλου και ταυτόχρονα δημιουργεί εικόνες με καλύτερη ανάλυση. Η αξιολόγηση της παραγόμενης εικόνας γίνεται με το να χρησιμοποιηθεί αυτή η εικόνα ως είσοδος σε μία αναζήτηση. Τα αποτελέσματα της αναζήτησης αυτής είναι εικόνες φαγητών που υπάρχουν ήδη. Φαίνεται λοιπόν ότι το αποτέλεσμα της αναζήτησης από εικόνα που έχει προέλθει από απλό GAN διαφέρει ενώ το αποτέλεσμα της αναζήτησης από εικόνα που έχει προέλθει από το WGAN-GP (εικόνα 1.5) είναι αρκετά παρόμοιο με την παραγόμενη εικόνα [16].

Query



Input ingredients
Tomato + Chicken

Search result



香りよし！トマトと紅茶の炊き込みご飯☆

紅茶の風味が香る、いな一品です！

Ingredients

- Tomato
- Chicken
- mushroom
- Rice
- Spice

Query



Input ingredients
Tomato + Chicken

Search result



鶏肉とトマトのてりやき！

トマトを入れることで、あっます。冷めても美味しくいた

Ingredients

- Eqq plant
- Green pepper
- Tomato
- Chicken
- salt-marinated rice malt

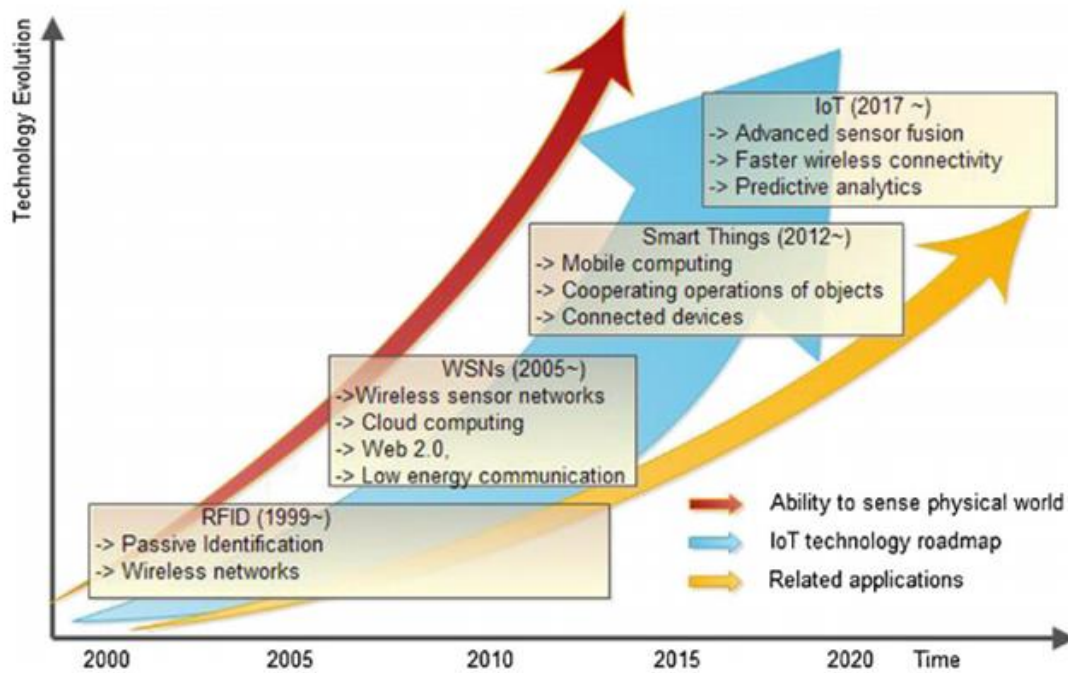
Εικόνα 1. 5: Αριστερά: Αξιολόγηση εικόνας που έχει προέλθει από απλό GAN. Η πάνω είναι η παραγόμενη εικόνα και η κάτω είναι η εικόνα μετά την αναζήτηση με "φίλτρο" την παραγόμενη εικόνα. Δεξιά: Αξιολόγηση εικόνας που έχει προέλθει από WGAN-GP. Η πάνω είναι η παραγόμενη εικόνα και η κάτω είναι η εικόνα μετά την αναζήτηση με "φίλτρο" την παραγόμενη εικόνα [16].

Κεφάλαιο 2: Θεωρητικό υπόβαθρο

2.1 Διαδίκτυο των πραγμάτων

Όπως ήδη αναφέρθηκε τα τελευταία χρόνια το διαδίκτυο έγινε αναπόσπαστο κομμάτι πολλών “πραγμάτων” και συσκευών, δημιουργώντας μία νέα κατηγορία συσκευών, τις “έξυπνες συσκευές”. Οι λεγόμενες “έξυπνες συσκευές” ανήκουν σε μία μεγαλύτερη ομάδα γνωστή και ως “Διαδίκτυο των πραγμάτων” (Internet of Things - IoT) που φαίνεται να προσελκύει την προσοχή των επιστημόνων. Το IoT παρουσιάζεται ως το μέλλον του διαδικτύου, με δισεκατομμύρια “έξυπνα αντικείμενα” να εκτιμάται ότι πρόκειται να δημιουργηθούν. Συγκεκριμένα η φιλοσοφία του IoT βασίζεται σε ένα δίκτυο από συνδεδεμένες συσκευές ή αντικείμενα, που αυτά θα συνδέονται ασύρματα μέσω αισθητήρων [1] [17]. Σύμφωνα με ένα άρθρο που δημοσιεύτηκε το 2015 υπήρχαν 9 δισεκατομμύρια διασυνδεδεμένες συσκευές και προέβλεπε την αύξηση αυτού του αριθμού στα 24 δισεκατομμύρια μέχρι το 2020. Ο Εθνικός Σύμβουλος Πληροφοριών των ΗΠΑ ισχυρίστηκε ότι “έως το 2025 ότι κόμβοι του διαδικτύου μπορεί να βρίσκονται σε καθημερινά πράγματα όπως το φαγητό, τα έπιπλα και άλλα” [18]. Επίσης το IoT μπορεί να αλληλοεπιδράσει χωρίς την ανθρώπινη παρέμβαση.

Διάφορες εφαρμογές του IoT έχουν ήδη κάνει τα πρώτα βήματα σε τομείς της υγείας, των μεταφορών αλλά και της αυτοκινητοβιομηχανίας. Παρόλο που η τεχνολογία που περιγράφεται βρίσκεται σε πρώιμο στάδιο, πολλές νέες εφαρμογές έχουν ενσωματώσει στα αντικείμενα κατάλληλους αισθητήρες που συνδέονται στο διαδίκτυο. Αξίζει να αναφερθεί ότι η πρώτη αναφορά στην ιδέα του Internet of Things έγινε το 1999 από τον Kevin Ashton [1]. Ανέφερε συγκεκριμένα σε ένα άρθρο του, πως είμαστε φυσικά όντα όπως και το περιβάλλον μας. Συνέχισε λέγοντας πως η οικονομία μας, η κοινωνία μας και η γενικότερη επιβίωση μας δεν βασίζονται σε ιδέες και πληροφορίες αλλά σε πράγματα. “Δεν μπορείς να φας ψηφία (bits)”, συμπληρώνοντας πως αν και οι ιδέες και οι πληροφορίες είναι σημαντικές, τα αντικείμενα έχουν μεγαλύτερη αξία [19]. Συνέδεσε τον όρο με αντικείμενα που είναι συνδεδεμένα μεταξύ τους και μπορούν να αναγνωριστούν και να γίνει χρήση αυτών για εξαγωγή διάφορων πληροφοριών μέσω ραδιοσυχνότητων (RFID) [1]. Όπως υποστήριξε και ο δημιουργός της φράσης Internet of Things οι ραδιοσυχνότητες (RFID) και οι αισθητήρες (sensor technology) επιτρέπουν στους υπολογιστές να παρατηρούν, να αναγνωρίζουν και να καταλαβαίνουν τον κόσμο μέσω καταχωρημένων δεδομένων προσπερνώντας ανθρώπινους περιορισμούς [19]. Αν και ο ορισμός του IoT είναι ακόμα υπό διερεύνηση, γενικά ορίζεται ως μια παγκόσμια δυναμική (dynamic) υποδομή δικτύου με δυνατότητες αυτόματης διαμόρφωσης, με βάση συγκεκριμένα πρότυπα και πρωτόκολλα επικοινωνίας. Επίσης τα φυσικά ή εικονικά “πράγματα” του IoT έχουν ταυτότητα, συγκεκριμένα χαρακτηριστικά και μπορούν να χρησιμοποιήσουν “έξυπνες διεπαφές” και να ενσωματωθούν σε ένα δίκτυο πληροφοριών. Το IoT θα μπορούσε να χαρακτηριστεί και ως ένα υπερσύνολο των συνδεδεμένων συσκευών οι οποίες αναγνωρίζονται μόνο από υπάρχουσες τεχνικές (near field communication techniques NFC). Οι λέξεις «Διαδίκτυο» και «Πράγματα» σημαίνουν ένα διασυνδεδεμένο παγκόσμιο δίκτυο που βασίζεται σε αισθητήρες, στην επικοινωνία, σε δίκτυα και τεχνολογίες επεξεργασίας πληροφοριών, οι οποίες ίσως αποτελούν την νέα εκδοχή της τεχνολογίας των επικοινωνιών και πληροφοριών (information and communications technology (ICT)). Μέχρι σήμερα πολλές τεχνολογίες όπως τα δίκτυα ασύρματων αισθητήρων (WSN), οι γραμμωτοί κώδικες (barcodes), η ευφυής αντίληψη (intelligent sensing), RFID, NFC, επικοινωνίες χαμηλής ενέργειας (low energy communications) και cloud computing πλαισιώνουν την ιδέα του IoT. Η ασύρματη επικοινωνία έχει γίνει αναπόσπαστο κομμάτι του IoT ανοίγοντας νέους ορίζοντες εξερεύνησης και ανάπτυξης, όπου “φυσικά αντικείμενα” θα αναγνωρίζονται και θα κατευθύνονται μέσω του Διαδικτύου [1]. “Το Internet of Things έχει την δυναμική να αλλάξει τον κόσμο, ίσως και σε μεγαλύτερο βαθμό από ότι ήδη έχει κάνει το Διαδίκτυο” προβλέπει ο Kevin Ashton [19] [20].



Εικόνα 2. 1: Η εξέλιξη του Internet of Things [1].

2.1.1 Αρχιτεκτονική του διαδικτύου των πραγμάτων

Η αρχιτεκτονική του Internet of Things είναι γνωστή και αποτελείται από τρία στρώματα, το "Perception Layer", το "Network Layer" και το "Application Layer" [21].

2.1.1.1 Το Perception Layer

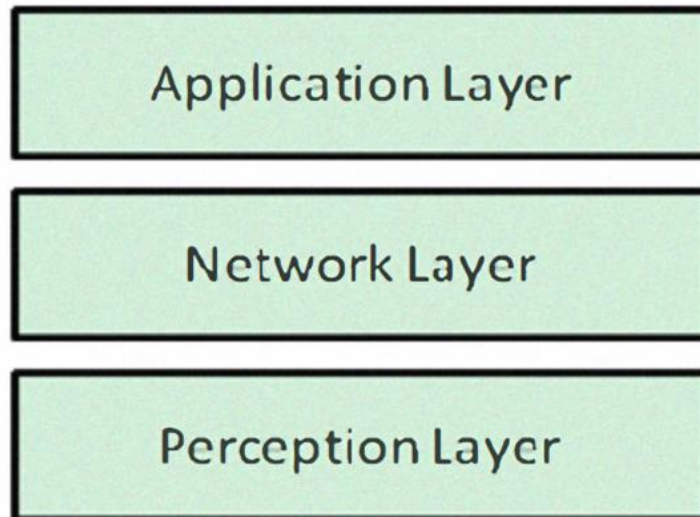
Το στρώμα αυτό της αρχιτεκτονικής του IoT είναι σαν το δέρμα του προσώπου και τις πέντε αισθήσεις του ανθρώπου δηλαδή έχει την ικανότητα να αναγνωρίζει τα αντικείμενα και να συλλέγει πληροφορίες. Έτσι και το IoT θέλοντας να επικοινωνεί με το περιβάλλον του περιλαμβάνει τα εξής: γραμμικό κώδικα δυο διαστάσεων (2-D bar code), ετικέτες RFID. Επίσης περιλαμβάνει κάμερα, GPS, αισθητήρες, τερματικά και δίκτυο αισθητήρων. Φυσικά έχει την δυνατότητα να αποκρυπτογραφήσει τις πληροφορίες που λαμβάνει. Ο βασικός του σκοπός είναι να αναγνωρίζει και να προσδιορίζει αντικείμενα αλλά και να συλλέγει πληροφορίες [21].

2.1.1.2 Το Network Layer

Το επίπεδο αυτό θα μπορούσε να παρομοιαστεί με το νευρικό σύστημα και τον εγκέφαλο του ανθρώπου. Η βασική λειτουργία αυτού του επιπέδου είναι η μετάδοση και η επεξεργασία των πληροφοριών. Αυτό το στάδιο περιλαμβάνει μια ένωση του δικτύου επικοινωνίας, του δικτύου του διαδικτύου, του δικτύου διαχείρισης, πληροφοριών και "έξυπνης" επεξεργασίας. Ουσιαστικά το Network Layer θα μεταδώσει και θα επεξεργαστεί τις πληροφορίες που παρέλαβε από το προηγούμενο επίπεδο (Perception Layer) [21].

2.1.1.3 To Application Layer

Το τελευταίο αυτό επίπεδο είναι η διεπαφή μεταξύ των συσκευών του διαδικτύου των πραγμάτων και του δικτύου με το οποίο επικοινωνούν. Με λίγα λόγια μέσω αυτού του επιπέδου μεταφέρονται τα δεδομένα και οι πληροφορίες ανάμεσα στις συσκευές του IoT και στο δίκτυο. Στο επίπεδο αυτό υλοποιείται η μορφοποίηση και η παρουσίαση των δεδομένων. Σε έναν υπολογιστή το application layer υλοποιείται μέσω των διαφόρων προγραμμάτων περιήγησης. [21].



Εικόνα 2. 2: Τα τρία επίπεδα της αρχιτεκτονικής του Internet of Things [21].

2.2 Έξυπνες συσκευές

Οι “έξυπνες συσκευές” (smart things) είναι μια ομάδα συσκευών που μπορούν να παρακολουθούνται και να ελέγχονται μέσω άλλων συσκευών (hub device) και υπηρεσιών διαδικτύου (web services). Οι χρήστες των “έξυπνων συσκευών” θα μπορούν να ελέγχουν και να αυτοματοποιούν καθημερινές δραστηριότητες μέσω αυτών των συσκευών [17]. Ο “έξυπνος” αυτός κόσμος περιλαμβάνει ψυγεία, “έξυπνα πλυντήρια”, “έξυπνη τηλεόραση”, “έξυπνες συσκευές” σπιτιού, “έξυπνα παπούτσια” και “έξυπνα τηλέφωνα” που χρησιμοποιούνται ήδη. Ωστόσο η εξοικείωση των χρηστών σε αυτή την ιδέα είναι ακόμα σε πρώιμο στάδιο. Η μετάβαση του παγκόσμιου διαδικτύου στο IPv6 θα προσφέρει απεριόριστο αριθμό δημόσιων διευθύνσεων IP, δημιουργώντας έτσι αμφίδρομες κι συμμετρικές συνδέσεις μεταξύ δισεκατομμυρίων “έξυπνων συσκευών”. Ειδικότερα το IPv6 είναι η έκτη έκδοση του Internet Protocol, που είναι η πιο πρόσφατη έκδοση του πρωτοκόλλου του διαδικτύου, δηλαδή του βασικού πρωτοκόλλου επικοινωνίας πάνω στο οποίο έχει χτιστεί όλο το διαδίκτυο [22]. Επομένως η γεφύρωση όλων αυτών των συσκευών θα πραγματοποιηθεί με την βοήθεια του IPv6 αλλά και την ενσωμάτωση των IoT με το Cloud. Ο όρος cloud χρησιμοποιείται γενικά για την περιγραφή κέντρων δεδομένων που χρησιμοποιούνται από πολλούς χρήστες μέσω του διαδικτύου. Ουσιαστικά μιλάμε για χώρους αποθήκευσης δεδομένων και υπολογιστικής ισχύος, χωρίς απαραίτητα άμεση και ενεργή διαχείριση από τον χρήστη [23] [18].

Οι “έξυπνες συσκευές” οι οποίες είναι αυτόνομες έχουν άμεση πρόσβαση και αλληλεπίδραση με ένα μεγαλύτερο δίκτυο. Η εισαγωγή της ευφυΐας στα αντικείμενα τα καθιστά αυτόματα πιο ασφαλή, βολικά και αποδοτικά συμβάλλοντας στην προσπάθεια για αυτοματοποίηση πολλών εργασιών αλλά και στην χρήση τους από απομακρυσμένα μέρη. Η “έξυπνάδα” των συσκευών αυτών μπορεί να προέλθει από διαφορετικές ενσωματωμένες τεχνολογίες του Internet of Things όπως το RFID, barcodes, IPv4/IPv6, αισθητήρες, GPS, Wi-fi, Bluetooth και άλλα. Συγκεκριμένα ονομάζουμε “έξυπνο” οποιοδήποτε φυσικό αντικείμενο μπορεί να συνδεθεί στο διαδίκτυο και να έχει κάποιες δυνατότητες αντίληψης. Ειδικότερα αυτές οι δυνατότητες είναι αρχικά να μπορεί να εντοπίζει τους χρήστες και τις κοινωνικές συνδέσεις μεταξύ τους αλλά και να έχει πρόσβαση στα δεδομένα του χρήστη. Τέλος βασικές δυνατότητες μιας “έξυπνης συσκευής” είναι να συντονίζει τις λειτουργίες της αλλά και να παρέχει μία έξοδο ανάλογα με την λειτουργία που εκτελεί. Άλλος

ένας ορισμός που μπορούμε να δώσουμε στα “έξυπνες συσκευές” είναι ότι χαρακτηρίζονται από την ικανότητά τους να εκτελούν πολλαπλές, ίσως και ταυτόχρονες, εφαρμογές, υποστηρίζοντας κάθε φορά διαφορετικό βαθμό προσαρμογής. Επιπροσθέτως είναι εφικτό να χρησιμοποιηθούν από τον χρήστη από απόσταση, σύμφωνα φυσικά με τους τοπικούς περιορισμούς πόρων που υπάρχουν. Οι “έξυπνες συσκευές” λειτουργούν υπό τον έλεγχο των ανθρώπων όπως οι προσωπικοί υπολογιστές, τα κινητά τηλέφωνα, οι κάμερες, οι κονσόλες παιχνιδιών και άλλα. Οι δημιουργοί του IoT προβλέπουν ότι στο μέλλον οι “έξυπνες συσκευές” θα μπορούν να κατανοήσουν το περιβάλλον τους και να αντιδράσουν με βάση αυτό, χωρίς την ανθρώπινη παρέμβαση [18].

Η ιδέα των “έξυπνων συσκευών” έχει τέσσερα λογικά αρχιτεκτονικά επίπεδα ξεκινώντας από το πρώτο που είναι να συνδεθεί η συσκευή με το “smart things hub” ή σε μερικές περιπτώσεις κατευθείαν με το cloud. Το smart things hub είναι μία συσκευή που πάνω σε αυτή συνδέονται άλλες “έξυπνες συσκευές” ώστε να επικοινωνούν μεταξύ τους. Έπειτα το “smart hub” λειτουργεί σαν πύλη ώστε να ληφθούν/προωθηθούν μηνύματα, πληροφορίες και συμβάντα από και προς το cloud. Το τρίτο στάδιο περιέχει τις υπηρεσίες του διαδικτύου που γίνονται διαθέσιμες στις “έξυπνες συσκευές”. Τέλος το τέταρτο και τελευταίο στάδιο είναι η παρουσίαση με την μορφή των εφαρμογών. Βέβαια υπάρχουν άλλα τέσσερα επίπεδα αρχιτεκτονικής ενσωματωμένα στο cloud. Πρώτα από όλα η “συνδεσιμότητα”, η οποία είναι υπεύθυνη για την διατήρηση της μόνιμης σύνδεσης μεταξύ των smart things hub και των smart things mobile application δηλαδή των εφαρμογών των “έξυπνων συσκευών”. Έπειτα γίνεται η “επεξεργασία και η δρομολόγηση των συμβάντων” από συσκευές ή hub προς “έξυπνες εφαρμογές” που σχετίζονται με συγκεκριμένες συσκευές ή συμβάντα. Στο τρίτο στάδιο (“εφαρμογής”) παρέχεται η πρόσβαση στα δεδομένα, για δεδομένα σχετικά με λογαριασμούς και χρήστες που χρησιμοποιούνται από τις “έξυπνες εφαρμογές”. Ως τελευταίο στάδιο έχουμε τις “υπηρεσίες του διαδικτύου” οι οποίες παρέχονται όπως και το Application Programming Interface (API) [18]. Το API είναι ένα λογισμικό που επιτρέπει σε δύο εφαρμογές να επικοινωνούν μεταξύ τους, ανταλλάσσοντας δεδομένα και πληροφορίες εύκολα και ασφαλές [24].

2.2.1 Διαχωρισμός των έξυπνων συσκευών

Όπως αναφέρθηκε και προηγουμένως η τεχνολογία των “έξυπνων συσκευών” έχει αρχίσει να ανθίζει και να εξελίσσεται όλο και περισσότερο. Όλο και περισσότερα “ευφυή αντικείμενα” δημιουργούνται μεγαλώνοντας τον κόσμο τους. Αυτό δημιουργεί την ανάγκη για κατηγοριοποίηση όλων αυτών των αντικειμένων ώστε να είναι ευκολότερο να κατανοηθούν οι διαφορές τους και να αναγνωριστούν οι δυνατότητες τους. Ο διαχωρισμός των αντικειμένων έγινε βάση των δυνατοτήτων τους, δηλαδή τα “ευφυή αντικείμενα” με περίπλοκες και μεγαλύτερες ικανότητες κατηγοριοποιούνται σε διαφορετική κατηγορία από αυτή των αντικειμένων με περιορισμένες δυνατότητες. Σύμφωνα με διάφορες εργασίες που έχουν πραγματοποιηθεί, πολλοί έχουν προσπαθήσει να δημιουργήσουν ένα μοντέλο ταξινόμησης. Το πώς θα γίνει βέβαια η κατηγοριοποίηση αυτή, ποικίλει ανάλογα με τις ιδιότητες και τα χαρακτηριστικά που μελετώνται. Ωστόσο μπορούν να διακριθούν σε δύο κατηγορίες. Η πρώτη είναι η “nesting oriented” η οποία προτείνει έναν βασικό τύπο όπου έπειτα υπάρχουν πιο σύνθετες τάξεις. Η δεύτερη είναι η “non-nesting oriented” όπου σε αυτήν την κατηγορία δεν έχουμε έναν βασικό τύπο που μπορεί να επεκταθεί και να έχει εν τέλει πιο σύνθετες ιδιότητες. Όπως ίσως γίνεται αντιληπτό και από την μετάφραση αυτών των δύο κατηγοριών στην πρώτη κατηγορία ο βασικός τύπος επεκτείνεται και γίνεται η βάση για να “φωλιάσουν” άλλες πιο περίπλοκες ιδιότητες, μιλώντας σε θεωρητικό επίπεδο, ενώ στην δεύτερη κατηγορία δεν ισχύει αυτό.

Μια άλλη ιδέα για την κατηγοριοποίηση των “έξυπνων συσκευών” είναι με βάση τον σχεδιασμό τους. Συγκεκριμένα μέσω αυτής της κατηγοριοποίησης προκύπτουν τρεις τύποι αντικειμένων, δηλαδή τρεις κατηγορίες. Η “Activity-aware objects”, η “Policy-aware objects” και η “Process-aware objects”. Η πρώτη κατηγορία, η “Activity-aware objects”, σχετίζεται με τις ενέργειες που εκτελούνται από τα “έξυπνα αντικείμενα” και γενικότερα με το πλαίσιο που ακολουθούν. Η δεύτερη κατηγορία, η “Policy-aware objects”, ορίζει τον τρόπο προγραμματισμού των αντικειμένων και σχετίζεται με τα βασικά πρωτόκολλα που εκδίδονται είτε από την κυβέρνηση είτε από την διαχειριστή του κάθε συστήματος. Η τελευταία κατηγορία είναι η “Process-aware objects” που έχει να κάνει με την σειρά των ενεργειών και κινήσεων που εκτελούνται, δηλαδή με την αλληλεπίδραση μεταξύ των “έξυπνων συσκευών” [25].

2.3 Όραση υπολογιστών

Η όραση υπολογιστών (computer vision) έχει ως κεντρικό άξονα την επίτευξη εξαγωγής πληροφοριών από δεδομένα εικόνας. Σκοπός της λοιπόν είναι να καταφέρει να “κατανοήσει” εικόνες όπως θα το έκανε και ένας ανθρώπινος νους. Η όραση των υπολογιστών είναι ένας πολύ μεγάλος κλάδος που περιλαμβάνει μεγάλη γκάμα εφαρμογών που σχετίζονται με την επεξεργασία οπτικών πληροφοριών. Θα μπορούσαμε να διακρίνουμε τρεις βασικούς στόχους που επιδιώκει να καταφέρει το computer vision [26].

Ο πρώτος βασικός στόχος είναι να καταφέρει να αναγνωρίζει αντικείμενα. Ειδικότερα το αποτέλεσμα που επιδιώκεται είναι ένας αλγόριθμος κατανόησης να καταφέρει να αναγνωρίζει αντικείμενα με την ίδια ακρίβεια που θα το έκανε και ένας άνθρωπος. Οι αλγόριθμοι καταφέρνουν να ξεχωρίσουν αντικείμενα σε μια φωτογραφία μέσω της έντασης στα διάφορα σημεία της εικόνας. Ειδικότερα ένα είδος διαχωρισμού γίνεται όταν περιοχές της εικόνας έχουν συγκεκριμένες περιοχές ομοιόμορφης έντασης ή μέσω των καμπυλών τοπικών μεγίστων ή ελαχίστων της έντασης. Με λίγα λόγια η αναγνώριση αντικειμένων γίνεται μέσω της μέτρησης της έντασης υπό διάφορες γωνίες αλλά και φίλτρα που καθιστούν δυνατό τον εντοπισμό της. Φυσικά μετ’ έπειτα η εκάστοτε εικόνα υπόκειται σε περαιτέρω επεξεργασία. Επίσης μια περιοχή εικόνας μπορεί να καθοριστεί/αναγνωριστεί μέσω στατιστικών παραμέτρων όπως η μέση ένταση και η τυπική απόκλιση. Επιπροσθέτως η αναγνώριση είναι εφικτό να γίνει μέσω γεωμετρικών ιδιοτήτων κάθε περιοχής όπως η αναλογία των διαστάσεων ενός σχήματος που περικλείει την περιοχή της εικόνας που εξετάζεται. Είναι σημαντικό να αναφερθεί ότι πολλές φορές οι ταξινομητές, δηλαδή οι αλγόριθμοι αναγνώρισης εικόνων, εκπαιδεύονται σε συγκεκριμένα αντικείμενα. Μπορεί δηλαδή ένας ταξινομητής να έχει εκπαιδευτεί για να αναγνωρίζει τους διάφορους τύπους αεροσκαφών [26].

Ο δεύτερος στόχος της υπολογιστικής όρασης είναι η “αυτόνομη οδήγηση”. Σίγουρα έχουμε ακούσει όλοι για οχήματα που μπορούν να λειτουργήσουν χωρίς να υπάρχει κάποιος άνθρωπος να τα οδηγεί. Στην περίπτωση αυτή η μηχανική μάθηση πρέπει να παρακολουθεί και να αναγνωρίζει συνεχώς τα όρια του δρόμου ώστε το όχημα να βρίσκεται πάντα στην σωστή θέση. Επομένως σε αυτή την κατηγορία το computer vision σκοπεύει να καταφέρει να αναγνωρίζει τα οποιαδήποτε όρια υπάρχουν και να καθοδηγεί σωστά. Ένας δεύτερος στόχος είναι να αποφεύγει τυχόν εμπόδια που θα συναντήσει. Στην περίπτωση του παραδείγματος με το αυτοκινούμενο όχημα ο αλγόριθμος πρέπει να αναγνωρίσει τρισδιάστατα εμπόδια άγνωστα έως τώρα για εκείνο και να τα αποφύγει. Αυτό που ουσιαστικά χρειάζεται είναι να δημιουργηθεί μια ακριβής περιγραφή του περιβάλλοντος που περιβάλλει το όχημα. Έως τώρα η αναγνώριση του περιβάλλοντος επιτυγχάνεται μέσω λείζερ, ραντάρ αλλά και άλλων [26].

Ο τρίτος στόχος της μηχανικής μάθησης σχετίζεται με τον δεύτερο και είναι η ανάκτηση ενός πλήρους και ακριβούς τρισδιάστατου μοντέλου ενός αντικειμένου. Έπειτα το μοντέλο αυτό μπορεί να χρησιμοποιηθεί σε διάφορες άλλες εφαρμογές όπως η αναγνώριση αντικειμένων και άλλα. Γενικότερα η προσομοίωση εικόνας χρησιμοποιείται σε πολλούς κλάδους όπως η ψυχαγωγία ή η στρατιωτική εκπαίδευση. Συγκεκριμένα στην στρατιωτική εκπαίδευση η παροχή τρισδιάστατων μοντέλων είναι σχεδόν απαραίτητη. Ο τρόπος που κατασκευάζονται αυτά τα μοντέλα είναι τοποθετώντας πολλά και διαφορετικά γεωμετρικά σχήματα σε πολλές όψεις του αντικειμένου ταυτόχρονα. Ο στόχος του computer vision είναι να εξελίξει όσο το δυνατόν περισσότερο γίνεται την διαδικασία της κατασκευής αυτού του μοντέλου ώστε να υλοποιείται με ελάχιστη ανθρώπινη παρέμβαση. Οι αυτοματοποιημένοι στερεοφωνικοί αλγόριθμοι (automated stereo algorithms) είναι μια καλή προσέγγιση για τον σκοπό που συζητάμε. Με αυτόν τον τρόπο διάφορες ενδείξεις όπως για παράδειγμα οι σκιές μπορούν να χρησιμοποιηθούν για να ενισχύσουν την ορθή δημιουργία του μοντέλου [26].

2.4 Τεχνητή Νοημοσύνη

Artificial Intelligence ή, όπως θα μπορούσε να μεταφραστεί στα ελληνικά, “Τεχνητή νοημοσύνη” είναι η επιστήμη η οποία δημιουργεί “έξυπνες” μηχανές. Ορίζεται ως ο τομέας της επιστήμης και της μηχανικής που μελετά την υπολογιστική ικανότητα, την κατανόηση των μηχανημάτων, την δημιουργία αυτών ώστε εν τέλει να αποκτούν “νοημοσύνη” [27]. Στοχεύει κυρίως στο να κατασκευάζει “έξυπνα προγράμματα” πάνω στους υπολογιστές. Θα μπορούσε κάποιος να το παρομοιάσει με την επιστήμη όπου οι υπολογιστές προσπαθούν να καταλάβουν, και να μιμηθούν ίσως, την ανθρώπινη νοημοσύνη, όμως δεν είναι ακριβώς αυτό αφού η τεχνητή νοημοσύνη δεν έχει βιολογικούς περιορισμούς [28]. Τι σημαίνει όμως ευφυΐα; Σύμφωνα με το “Webster’s New World Dictionary (1988), και για να αποσαφηνιστεί ο ακριβής ορισμός της συγκεκριμένης

λέξης, η ευφυΐα ορίζεται ως η ικανότητα κατανόησης και μάθησης από την εκάστοτε εμπειρία αλλά και η διατήρηση της υπάρχουσας γνώσης. Εκτός από την διανοητική ικανότητα το λεξικό στον ορισμό της λέξης περιλαμβάνει και την ικανότητα να αντιδρά γρήγορα, αποτελεσματικά και επιτυχώς στα ερεθίσματα ή σε μία καινούργια κατάσταση χρησιμοποιώντας με επιτυχία την λογική και καταφέροντας εν τέλει να λύσει οποιοδήποτε πρόβλημα. Το συγκεκριμένο λεξικό, στον τομέα της ψυχολογίας, ορίζει την ευφυΐα ως την ικανότητα κάποιου να καταφέρνει να υλοποιεί συγκεκριμένες δοκιμασίες [29]. Άρα εάν ζητηθεί από μια μηχανή να εκτελέσει υπολογισμούς που έχει εκπαιδευτεί πάνω σε αυτούς, μπορεί να δώσει εντυπωσιακά αποτελέσματα και έτσι να μπορέσει να χαρακτηριστεί ως “κάπως έξυπνη”. Αξίζει να αναφερθεί ότι η νοημοσύνη των υπολογιστών δεν είναι πάντα παρόμοια με αυτή των ανθρώπων [28]. Σίγουρα παρατηρώντας το σκεπτικό, τις μεθόδους και την νοημοσύνη των ανθρώπων και χρησιμοποιώντας αυτά, μπορούν να ληφθούν συμπεράσματα ώστε εν τέλει να εκπαιδευτούν οι μηχανές, μέχρι ένα σημείο. Βέβαια οι περισσότερες έρευνες πάνω στην τεχνητή νοημοσύνη αφορούν καθημερινά προβλήματα του κόσμου και όχι να μελετήσουν τον άνθρωπο και τα ζώα. Γι’ αυτό και η τεχνητή νοημοσύνη συνηθίζει να χρησιμοποιεί μεθόδους με περισσότερους υπολογισμούς από αυτούς που κάνουν συνήθως οι άνθρωποι στην καθημερινή τους ζωή. Είναι σημαντικό επίσης να μην ταυτιστεί η τεχνητή νοημοσύνη των υπολογιστών με αυτή των ανθρώπων όπως ήδη έχει τονιστεί και κατ’ επέκταση να μετρηθεί η νοημοσύνη μιας μηχανής με ένα τεστ IQ. Για παράδειγμα σε ένα τεστ IQ σε ένα παιδί μπορεί να ζητηθεί να απομνημονεύσει μια σειρά από ψηφία, κάτι που για έναν υπολογιστή είναι υπερβολικά απλό. Ωστόσο ορισμένα από τα προβλήματα των IQ τεστ μπορεί να αποτελέσουν πρόκληση για την τεχνητή νοημοσύνη. Επιπροσθέτως οι υπολογιστές στις μέρες μας έχουν μεγάλη ταχύτητα και μνήμη, όμως το ποιες λειτουργίες καταφέρνουν να εκτελέσουν και πόσο καλά εξαρτάται από την ποιότητα της εκπαίδευσης που έχουν λάβει αλλά και από τους προγραμματιστές που τα εκπαίδευσαν [28].

2.5 Μηχανική Μάθηση

Η Μηχανική Μάθηση (Machine learning – ML) είναι ένας κλάδος της Τεχνητής Νοημοσύνης (Artificial Intelligence - AI). Όπως και ο ανθρώπινος εγκέφαλος έτσι και οι αλγόριθμοι της μηχανικής μάθησης μαθαίνουν από το περιβάλλον τους [30], δηλαδή είναι κατασκευασμένοι με τέτοιο τρόπο ώστε να βελτιώνονται συνεχώς μέσα από τα λάθη τους και να μαθαίνουν μέσω της εμπειρίας [31]. Είναι ένας από τους πιο ταχέως αναπτυσσόμενους τεχνικούς τομείς στις μέρες μας. Για να προσδιοριστεί το πεδίο έρευνας του ML μπορούμε να πούμε ότι είναι η τομή της επιστήμης των υπολογιστών με την στατιστική [32]. Μέσω της στατιστικής εξάγονται τιμές, οι οποίες προκύπτουν από την εκπαίδευση που έχουν υποστεί οι υπολογιστές πάνω σε σύνολα δεδομένων που δέχονται ως είσοδο [31]. Η μηχανική μάθηση έχει εφαρμοστεί σε διάφορους κλάδους όπως η αναγνώριση προτύπων, η όραση υπολογιστών, σε διάφορες βίο-ιατρικές εφαρμογές, σε τομείς της ψυχαγωγίας, στην εκπαίδευση την ιατρική, το μάρκετινγκ αλλά και σε πολλά άλλα [30]. Η μηχανική μάθηση θα μπορούσε να χωριστεί σε δύο κατηγορίες εκμάθησης, την επιβλεπόμενη και την μη επιβλεπόμενη όπου και θα αναλυθούν παρακάτω [31].

2.5.1 Επιβλεπόμενη Μάθηση

Το βασικό χαρακτηριστικό αυτής της εκπαίδευσης είναι οι ετικέτες (labels) που εισάγονται στα σύνολα εκπαίδευσης. Συνήθως αυτές οι ετικέτες χρησιμοποιούνται για να χαρακτηρίσουν τις κλάσεις σε προβλήματα ταξινόμησης. Ειδικότερα οι αλγόριθμοι του supervised learning επάγουν μοντέλα από τέτοιου είδους σύνολα εκπαίδευσης ώστε να χρησιμοποιηθούν μελλοντικά για την ταξινόμηση άγνωστων δεδομένων [33]. Αυτή η τεχνική εκμάθησης είναι παρόμοια με την ανθρώπινη μάθηση από παλαιότερες εμπειρίες με σκοπό την δημιουργία μνήμης και γνώσεων όπου θα χρησιμοποιηθούν μελλοντικά [34].

Το supervised learning στην πραγματικότητα είναι η εκμάθηση που γίνεται με παραδείγματα. Στο δίκτυο που πρόκειται να εκπαιδευτεί εισάγονται δύο σύνολα δεδομένων, ένα σύνολο εκπαίδευσης και ένα σύνολο δοκιμών. Ο σκοπός είναι το δίκτυο που θα εκπαιδευτεί να μάθει από ένα σύνολο δεδομένων με ετικέτες ώστε να αναγνωρίζει και να ταξινομεί όσο το δυνατόν καλύτερα παραδείγματα που δεν έχουν ετικέτες. Τα παραδείγματα αυτά ορίζουν το σύνολο δοκιμών. Με λίγα λόγια το ζητούμενο είναι το δίκτυο να αναπτύξει έναν κανόνα σύμφωνα με τον οποίο θα κατηγοριοποιεί νέα παραδείγματα, αναλύοντας παραδείγματα που έχουν ήδη δοθεί σε αυτό μαζί με την αντίστοιχη ετικέτα. Ένα αντίστοιχο παράδειγμα θα ήταν να δοθούν σαν σύνολο εκπαίδευσης ένα σύνολο που αποτελείται από διάφορα λαχανικά και που κάθε

εικόνα έχει μια ετικέτα που ορίζει το είδος του λαχανικού της εικόνας. Αφού το δίκτυο εκπαιδευτεί με το παραπάνω σύνολο, στόχος είναι να αναπτύξει έναν κανόνα σύμφωνα με τον οποίο θα μπορεί να αναγνωρίσει σε ποια κατηγορία ανήκει μια νέα εικόνα που θα του δοθεί η οποία όμως δεν έχει κάποια ετικέτα [35] [31].

2.5.2 Εκμάθηση χωρίς επίβλεψη

Οι μέθοδοι μάθησης χωρίς επίβλεψη (Unsupervised learning) έχουν αρχίσει να γίνονται μέρα με την μέρα όλο και πιο επιτυχημένες. Αυτή η ανάγκη έχει προκύψει λόγω αυξημένων δεδομένων χωρίς αντίστοιχη ετικέτα αλλά και λόγω της προσπάθειας να βελτιωθούν τα υπάρχοντα μοντέλα και να ενισχυθεί η ήδη υπάρχουσα γνώση [36]. Σε αυτή την εκμάθηση τα σύνολα δεδομένων που χρησιμοποιούνται για την εκπαίδευση δεν έχουν αντίστοιχες ετικέτες. Σκοπός λοιπόν της unsupervised μάθησης είναι να αναγνωρίζει μοτίβα ανάμεσα σε διάφορα στοιχεία όταν δεν υπάρχει ήδη κάποιος διαχωρισμός σε αυτά όπως γίνεται με τα labels. Μια υποκατηγορία του unsupervised learning είναι τα παραγωγικά μοντέλα (generative models). Σκοπός αυτών των δικτύων είναι να παράγει νέα δεδομένα τα οποία φαίνονται αληθινά ενώ δεν είναι [31].

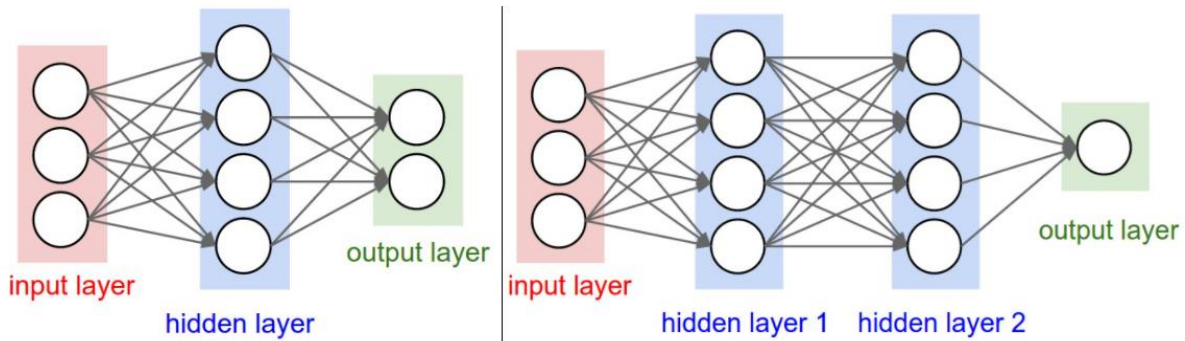
2.6 Βαθιά μάθηση

Η βάση του deep learning είναι τα νευρωνικά δίκτυα, τα οποία έπειτα από προσθήκη πολλαπλών κρυφών επιπέδων δημιουργούν τα βαθιά νευρωνικά δίκτυα. Η βαθιά μάθηση έχει συμβάλει στην αυτόματη επεξεργασία ήχου, εικόνας, γλώσσας, αναγνώριση φωνής, στην ταξινόμηση κειμένου αλλά και στην αναγνώριση και επεξεργασία διάφορων εικόνων και βίντεο [31]. Είναι μια κατηγορία της μηχανικής μάθησης η οποία δίνει την δυνατότητα στους υπολογιστές να επιλύουν προβλήματα, όπως η αναγνώριση εικόνας, ομιλίας. Αυτές οι μέθοδοι χρησιμοποιούν πολλαπλά επίπεδα επεξεργασίας ώστε να μαθαίνουν αναπαραστάσεις δεδομένων με πολλαπλά επίπεδα αφαίρεσης. Κάθε επίπεδο μαθαίνει μια έννοια από τα δεδομένα στα οποία βασίζονται τα επόμενα επίπεδα. Όσο πιο ψηλά βρίσκεται το επίπεδο τόσο πιο αφηρημένες είναι οι έννοιες που μαθαίνονται. Επίσης η βαθιά μάθηση δεν εξαρτάται από προηγούμενη επεξεργασία δεδομένων και εξάγει αυτόματα τα χαρακτηριστικά. Για παράδειγμα ένα βαθύ νευρωνικό δίκτυο πρέπει πρώτα να μάθει να αναγνωρίζει απλές ακμές σε διάφορα σχήματα και μετά να αναγνωρίζει ακμές σε πολύπλοκότερα σχήματα. Επιπλέον δεν υπάρχει κάποιος κανόνας για το πόσα επίπεδα απαιτούνται για να θεωρηθεί η μάθηση “βαθιά”, ωστόσο θεωρείται ότι απαιτούνται περισσότερα από δύο [37].

2.7 Τεχνητά Νευρωνικά Δίκτυα

Τα νευρωνικά δίκτυα μπορούν να διακριθούν στα βιολογικά και τα τεχνητά νευρωνικά δίκτυα (Artificial Neural Networks – ANN). Συγκεκριμένα τα βιολογικά είναι αυτά τα οποία υπάρχουν στο ανθρώπινο εγκέφαλο και είναι υπεύθυνα για τις λειτουργίες που διενεργούνται στον εγκέφαλο. Αποτελούνται από νευρώνες δηλαδή από τμήματα του εγκεφάλου που επεξεργάζονται ερεθίσματα και πληροφορίες. Αντλώντας έμπνευση από τα βιολογικά νευρωνικά δίκτυα, δημιουργήθηκαν τα τεχνητά νευρωνικά δίκτυα με σκοπό να λειτουργούν με παρόμοιο τρόπο με τα βιολογικά. Ουσιαστικά τα τεχνητά νευρωνικά δίκτυα είναι ένα μοντέλο επεξεργασίας δεδομένων. Ένα σύνολο σωστά συνδεδεμένων νευρώνων που επεξεργάζονται πληροφορίες αποτελούν ένα νευρωνικό δίκτυο. Επίσης επιμέρους ομάδες νευρώνων οργανώνονται σε στρώματα τα Layers [38]. Τα στρώματα αυτά συνδέονται σε μία σειρά με αρχή και τέλος. Από αυτό γίνεται αντιληπτό ότι οι έξοδοι μερικών νευρώνων μπορούν να αποτελέσουν είσοδο για κάποιους άλλους νευρώνες. Προφανώς υπάρχει ένα αρχικό και ένα τελικό στρώμα αφού δεν μπορούν να έχουν κυκλική μορφή τα νευρωνικά δίκτυα αφού κάτι τέτοιο θα σήμαινε έναν άπειρο βρόγχο χωρίς αποτέλεσμα. Αρχικά αποτελείται από το επίπεδο εισόδου (input layer) από το οποίο εισάγονται τα δεδομένα εισόδου. Στο επίπεδο αυτό δεν γίνεται κάποια επεξεργασία, απλά τροφοδοτεί το δίκτυο με τις πληροφορίες από το εξωτερικό περιβάλλον. Μετά το πρώτο επίπεδο τα τεχνητά νευρωνικά δίκτυα αποτελούνται από τα κρυφά επίπεδα (hidden layer/s).

Τα κρυφά επίπεδα μπορεί να είναι είτε ένα είτε περισσότερα και εκεί επεξεργάζονται τα δεδομένα που εισήχθησαν στο επίπεδο εισόδου. Όσο προχωράει η επεξεργασία στα κρυφά επίπεδα τόσο σημαντικότερες πληροφορίες λαμβάνονται από την επεξεργασία που γίνεται. Τέλος το τελευταίο επίπεδο είναι το επίπεδο εξόδου (output layer) στο οποίο το μοντέλο αποφασίζει το αποτέλεσμα που θα εξάγει σαν απάντηση, λαμβάνοντας υπόψη την ανάλυση και επεξεργασία των δεδομένων που έχει προηγηθεί. Στα κανονικά νευρωνικά δίκτυα (regular neural networks) ο πιο κοινός τύπος στρώματος είναι το πλήρως συνδεδεμένο στρώμα (fully connected layer) κατά το οποίο οι νευρώνες μεταξύ δύο γειτονικών στρωμάτων συνδέονται πλήρως κατά ζεύγη αλλά οι νευρώνες που βρίσκονται στο ίδιο στρώμα δεν συνδέονται μεταξύ τους [39]. Παρακάτω παρουσιάζονται δύο παραδείγματα τέτοιων νευρωνικών δικτύων.



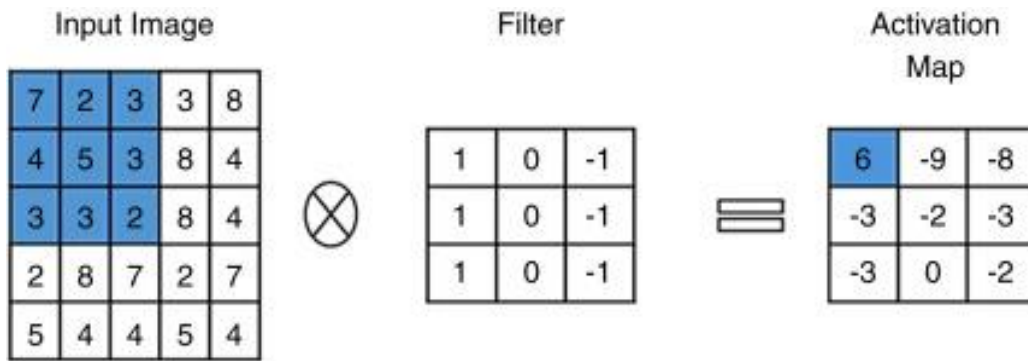
Εικόνα 2. 3 : Αριστερά απεικονίζεται ένα τεχνητό νευρωνικό δίκτυο 2 επιπέδων το οποίο αποτελείται από ένα επίπεδο εισόδου τριών εισόδων, από ένα κρυφό επίπεδο τεσσάρων νευρώνων και από ένα επίπεδο εξόδου δύο νευρώνων. Δεξιά απεικονίζεται ένα τεχνητό νευρωνικό δίκτυο 3 επιπέδων το οποίο αποτελείται από ένα επίπεδο εισόδου τριών εισόδων, από δύο κρυφά επίπεδα τεσσάρων νευρώνων το καθένα και από ένα επίπεδο εξόδου ενός νευρώνα. Αξίζει να παρατηρηθεί ότι και στα δύο παραδείγματα νευρωνικών δικτύων υπάρχουν συνάξεις, δηλαδή συνδέσεις, ανάμεσα στους νευρώνες που βρίσκονται σε διαφορετικό επίπεδο και όχι στο ίδιο. Η εικόνα πάρθηκε από [39].

2.8 Συνελικτικά Νευρωνικά δίκτυα

Τα CNN δίκτυα αποτελούνται από τρεις τύπους επιπέδων. Αυτά είναι τα συνελικτικά στρώματα (convolutional layers), τα στρώματα συγκέντρωσης (pooling layers) και τα πλήρως συνδεδεμένα στρώματα (fully connected layers).

2.8.1 Συνελικτικό επίπεδο

Τα συνελικτικά στρώματα (Convolutional layers) είναι το κύριο δομικό στοιχείο ενός convolutional neural network και άρα είναι απαραίτητα για την λειτουργία των CNN. Περιέχει ένα σύνολο φίλτρων ή πυρήνων (kernels) όπου οι παράμετροι αυτών πρέπει να εκπαιδευτούν κατά την διάρκεια της εκπαίδευσης. Το μέγεθος των φίλτρων είναι συνήθως μικρότερο από την εικόνα. Κάθε φίλτρο συνδυάζεται με την εικόνα ώστε να δημιουργηθεί ένας χάρτης ενεργοποίησης (activation map). Για την συνέλιξη το φίλτρο μετακινείται σε όλο το ύψος και πλάτος της εικόνας έτσι ώστε να υπολογιστεί το γινόμενο κάθε στοιχείου του φίλτρου με την είσοδο της εικόνας σε κάθε χωρική θέση. Η πρώτη καταχώρηση στον χάρτη ενεργοποίησης γίνεται περιστρέφοντας το φίλτρο με το μπλε χρώμα στην εικόνα εισόδου όπως φαίνεται στην εικόνα 2.4. Ο χάρτης ενεργοποίησης δημιουργείται αφού προσπελαστεί κάθε στοιχείο της εικόνας. Η έξοδος του συνελικτικού στρώματος δημιουργείται από την ένωση των χαρτών ενεργοποίησης κάθε φίλτρου. Συγκεκριμένα κάθε στοιχείο του χάρτη ενεργοποίησης μπορεί να θεωρηθεί ως η έξοδος ενός νευρώνα. Επίσης τα αρχικά συνελικτικά στρώματα εξάγουν τα “ευκολότερα” χαρακτηριστικά της εικόνας όπως για παράδειγμα γραμμές, ενώ τα τελευταία στρώματα εξάγουν δυσκολότερες πληροφορίες όπως σχήματα [40]. Με αυτή την διαδικασία, το δίκτυο θα εκπαιδεύσει τα φίλτρα ή τους kernels να “ενημερώνουν” όταν παρατηρήσουν ένα συγκεκριμένο χαρακτηριστικό σε μια συγκεκριμένη χωρική θέση της εισόδου [41].



Εικόνα 2. 4: Μια αναπαράσταση της συνέλιξης [40].

2.8.2 Στρώμα συγκέντρωσης

Το pooling layer ή αλλιώς το στρώμα συγκέντρωσης έχει ως στόχο να μειώσει τον αριθμό των παραμέτρων ώστε να μειωθεί και η υπολογιστική πολυπλοκότητα του δικτύου χρησιμοποιώντας την συνάρτηση “max”. Στα περισσότερα CNN δίκτυα, τα στρώματα αυτά έχουν την μορφή στρωμάτων max-pooling με kernels διαστάσεων 2x2. Γενικά τα στρώματα συγκέντρωσης αποτελούνται από pooling νευρώνες που μπορούν να εκτελέσουν πολλές λειτουργίες όπως η κανονικοποίηση L1/L2 και η μέση συγκέντρωση (average pooling) [41].

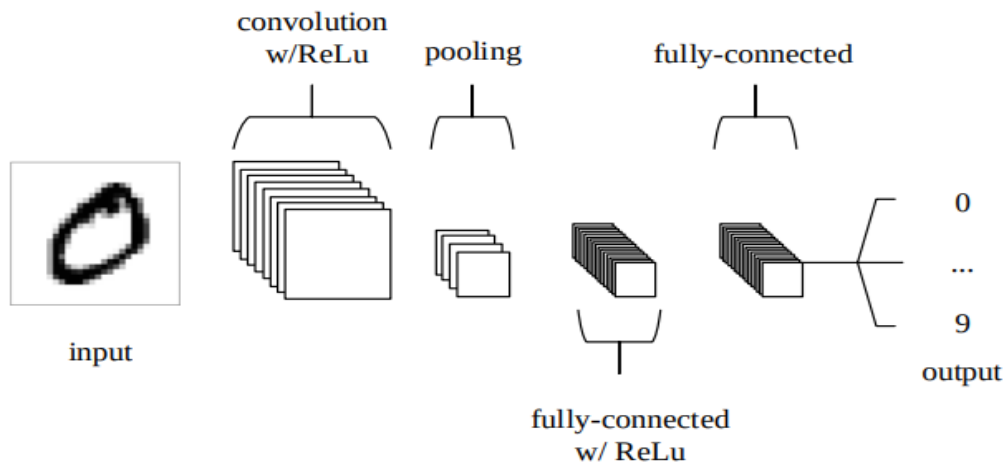
2.8.3 Πλήρως συνδεδεμένο επίπεδο

Τα πλήρως συνδεδεμένα στρώματα (Fully-connected layers) περιέχουν νευρώνες οι οποίοι συνδέονται άμεσα με τους γειτονικούς νευρώνες των γειτονικών στιβάδων, χωρίς ενδιάμεσα να συνδέονται με κάποιο άλλο στρώμα [41]. Είναι μία απλή λογική και όσων αναφορά το υπολογιστικό κομμάτι και τις πληροφορίες για όλη την εικόνα που μπορούμε να πάρουμε. Ωστόσο αφορά κυρίως μικρότερες εικόνες (8x8 pixels και 28x28 pixels) ενώ για μεγαλύτερες εικόνες η τεχνική αυτή αρχίζει να δυσκολεύει [42].

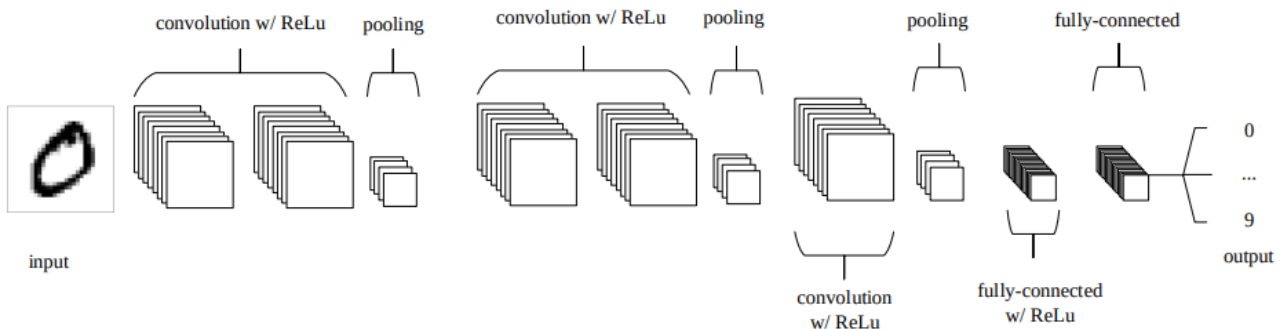
2.8.4 Συνελικτικό νευρωνικό δίκτυο για το σύνολο δεδομένων MNIST

Στην εικόνα 2.5 φαίνεται μια απλή αρχιτεκτονική ενός CNN δικτύου για το σύνολο δεδομένων MNIST. Αρχικά το δίκτυο του παραδείγματος της εικόνας 2.5 στο επίπεδο εισόδου του, κρατάει τις τιμές των pixels της εικόνας. Έπειτα το συνελικτικό στρώμα καθορίζει την έξοδο των νευρώνων. Μετ’ έπειτα το επίπεδο συγκέντρωσης θα εκτελέσει μια δειγματοληψία κατά μήκος της χωρικής διάστασης της εκάστοτε εισόδου. Έτσι θα μειωθεί κι άλλο ο αριθμός των παραμέτρων. Τέλος τα πλήρως συνδεδεμένα επίπεδα θα προσπαθήσουν να παράγουν διάφορα score για τις κλάσεις ώστε να χρησιμοποιηθούν τα score αυτά στην διαδικασία ταξινόμησης. Σε αυτά τα layers μπορεί να χρησιμοποιηθεί και το ReLu ώστε να βελτιωθεί η απόδοση του μοντέλου. Η συνάρτηση ReLu είναι μια τμηματικά γραμμική συνάρτηση που εξάγει κατ’ ευθείαν την είσοδο στην περίπτωση που είναι θετική, ενώ σε κάθε άλλη περίπτωση η έξοδος της είναι μηδέν. Χρησιμοποιείται σε πολλούς τύπους νευρωνικών δικτύων διότι τα μοντέλα που την χρησιμοποιούν εκτός του ότι επιτυγχάνουν μεγαλύτερη απόδοση, εκπαιδεύονται και ευκολότερα [43]. Μέσω της διαδικασίας που μόλις περιεγράφηκε, τα δίκτυα αυτά καταφέρνουν να μετασχηματίσουν το αρχικό επίπεδο εισόδου χρησιμοποιώντας τεχνικές συνελικτικής δειγματοληψίας ώστε να δημιουργήσουν τις απαραίτητες βαθμολογίες των κλάσεων και να εκτελεστεί τελικώς η ταξινόμηση. Φυσικά αυτή η αρχιτεκτονική που περιεγράφηκε πολλές φορές χρειάζεται

αρκετή βελτιστοποίηση. Άλλη μια κοινή αρχιτεκτονική που ακολουθούν τα δίκτυα αυτά είναι η ακολουθία δύο συνελκτικών στρωμάτων πριν από κάθε pooling layer όπως φαίνεται στην εικόνα 2.6 [41].



Εικόνα 2. 5: Μια απλή αρχιτεκτονική ενός συνελκτικού νευρωνικού δικτύου [41].



Εικόνα 2. 6: Μια αρχιτεκτονική ενός συνελκτικού νευρωνικού δικτύου με πολυπλοκότερη δομή [41].

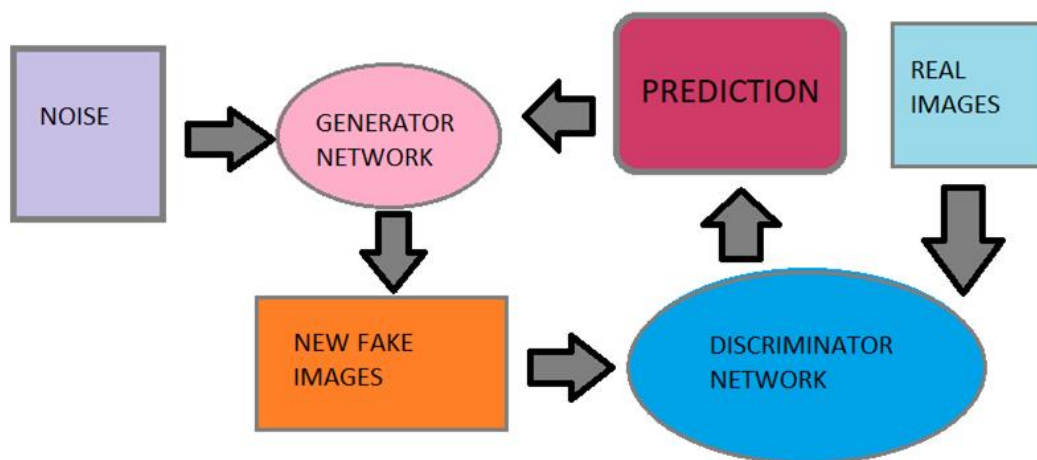
2.9 Παραγωγικά Ανταγωνιστικά Δίκτυα

Τα Παραγωγικά Ανταγωνιστικά Δίκτυα (Generative Adversarial Neural Networks - GANs) είναι μια αναδυόμενη τεχνολογία η οποία χρησιμοποιείται τόσο στην ημειπιβλεπόμενη μάθηση όσο και στην μη επιβλεπόμενη μάθηση. Τα παραγωγικά ανταγωνιστικά δίκτυα μπορούν να συγκριθούν και με τα δίκτυα διευκρίνησης, τα οποία δέχονται ως είσοδο κάποια δεδομένα και τα κατατάσσουν σε κάποια κατηγορία. Ουσιαστικά κατηγοριοποιούν δεδομένα με παρόμοια χαρακτηριστικά χρησιμοποιώντας την πιθανότητα $p(y|x)$. Θα μπορούσαν λοιπόν να θεωρηθούν τα GANs ότι είναι το ανάποδο από τα δίκτυα διευκρίνησης αφού σκοπός των παραγωγικών ανταγωνιστών δικτύων είναι να δημιουργούν “έξοδο” με συγκεκριμένα χαρακτηριστικά, δηλαδή που θα ανήκουν σε μία συγκεκριμένη κατηγορία [44]. Αυτή η τεχνολογία των παραγωγικών μοντέλων προτάθηκε το 2014 από τον Goodfellow [44] και αποτελείται από ένα ζευγάρι νευρωνικών δικτύων που λειτουργούν ανταγωνιστικά το ένα στο άλλο και επιτελεί το καθένα μία διαφορετική λειτουργία. Θα μπορούσαμε να παρομοιάσουμε το ένα νευρωνικό ως τον “πλαστογράφο”, που σκοπό έχει να δημιουργεί πλαστογραφίες κάνοντας τες να φαίνονται ρεαλιστικές εικόνες. Ο “πλαστογράφος” είναι γνωστός στην βιβλιογραφία των GAN και ως “δημιουργός” G (generator) και δημιουργεί πλαστά δείγματα από διανύσματα τυχαίου θορύβου [45]. Το δεύτερο νευρωνικό θα χαρακτηριζόταν ως “διευκρινιστής” D (discriminator) ο οποίος παρέχει ανατροφοδότηση για την πιθανότητα σχετικά με το πόσο καλά τα παραγόμενα δείγματα μοιάζουν σε μία πραγματική κατανομή [45]. Σκοπός του δηλαδή είναι να λαμβάνει πλαστές εικόνες και

πραγματικές και να ξεχωρίζει τις πλαστές (αυτές δηλαδή που προέρχονται από τον δημιουργό) από τις αυθεντικές (αυτές που έχουν προέλθει από το σύνολο των πραγματικών δεδομένων). Αυτά τα δύο νευρωνικά δίκτυα εκπαιδεύονται παράλληλα και το ένα ανταγωνίζεται το άλλο. Τα GAN έχουν χρησιμοποιηθεί με επιτυχία για την δημιουργία εικόνας γενικότερα, εικόνας από κείμενο, ανακατασκευής εικόνων που δεν είναι πλήρεις ώστε να φαίνονται φυσικές (image in-painting) και ανάλυσης εικόνων χαμηλής ανάλυσης σε υψηλότερη (image super-resolution) [45]. Τα δίκτυα του δημιουργού G και του διευκρινιστή D είναι δίκτυα πολλαπλών επιπέδων που αποτελούνται από συνελκτικά δίκτυα ή πλήρως συνδεδεμένα επίπεδα βαθιών δικτύων ή και τα δύο μαζί. Συγκεκριμένα ο δημιουργός υπερδειγματοληπτει το διάνυσμα τυχαίου θορύβου που δέχεται μέχρι να δημιουργήσει μία εικόνα στο μέγεθος που θέλει. Αντίθετα ο διευκρινιστής είναι ένα αντίστροφο συνελκτικό δίκτυο, που υποδειγματοληπτει την εικόνα που παίρνει ως είσοδο και δίνει ως έξοδο την κατάλληλη πιθανότητα [44].

2.9.1 Λειτουργία Παραγωγικών Ανταγωνιστικών Δικτύων

Ο δημιουργός G δεν έχει πρόσβαση σε πραγματικές εικόνες, μόνο από την αλληλεπίδραση του με τον διευκρινιστή D μπορεί να διδαχθεί. Ειδικότερα ο δημιουργός παίρνει σαν είσοδο τυχαίους αριθμούς και κατασκευάζει μία εικόνα ως έξοδο [44]. Έπειτα αυτές οι συνθετικές εικόνες του δημιουργού τροφοδοτούνται στον διευκρινιστή που έχει πρόσβαση και στις πραγματικές εικόνες που υπάρχουν σε κάποια πιθανή βάση δεδομένων. Εάν ο δ διευκρινιστής αντιληφθεί ότι η εικόνα που αξιολογεί προέρχεται από τον δημιουργό G και όχι από την στοίβα με τις πραγματικές εικόνες στέλνει μήνυμα λάθους. Αυτό το σήμα του σφάλματος μπορεί να χρησιμοποιηθεί από τον δημιουργό για την εκπαίδευση του, ώστε να δημιουργήσει στο μέλλον μια καλύτερης ποιότητας πλαστογραφία/εικόνα. Σε ένα βασικό GAN ο discriminator αφού ελέγξει την εκάστοτε εικόνα, την βαθμολογεί με έναν αριθμό από το 0 έως το 1. Όσο πιο κοντά στο 0 είναι ο αριθμός αυτός, τόσο πιο πιθανό σύμφωνα με τον "διευκρινιστή" να έχει δημιουργηθεί αυτή η εικόνα από τον δημιουργό. Από την άλλη πλευρά όσο πιο κοντά στο 1 ο αριθμός που βάζει, τόσο πιο πιθανό να είναι πραγματική εικόνα η συγκεκριμένη, σύμφωνα πάντα με τον διευκρινιστή (discriminator). Με λίγα λόγια λοιπόν έχουμε διπλή ανατροφοδότηση αφού ο δημιουργός τροφοδοτεί τον διευκρινιστή με δικές του "πλαστές" εικόνες αλλά ανατροφοδοτείται από εκείνον με το αποτέλεσμα του ώστε να εκπαιδευτεί και να δημιουργήσει μελλοντικά μία καλύτερα εικόνα με σκοπό να τον ξεγελάσει. Αντίστοιχα ο διευκρινιστής ανατροφοδοτείται με αυθεντικές εικόνες ώστε να βελτιωθεί και να μην μπορεί να ξεγελαστεί από τον δημιουργό. Τα δίκτυα αυτά του GAN, δηλαδή του δημιουργού και του διευκρινιστή έχουν ως σκοπό να ελαχιστοποιήσουν την δικιά τους συνάρτηση κόστους [44]. Μία διαφορά που εντοπίζεται ανάμεσα σε αυτά τα δύο ανταγωνιστικά δίκτυα είναι ότι όταν ο διευκρινιστής γίνει βέλτιστος δε μπορεί να βελτιωθεί περαιτέρω ενώ ο δημιουργός συνεχίζει την εκπαίδευσή του μειώνοντας περισσότερο την ακρίβεια του διευκρινιστή. Συγκεκριμένα εάν ο generator καταφέρει να φτιάχνει εικόνες που δεν αποκλίνουν από τις αυθεντικές/ρεαλιστικές, ο discriminator μη μπορώντας να διαχωρίσει τις εικόνες δίνει τιμή 0,5 σε όλες τις εικόνες που έχει ως είσοδο.



Εικόνα 2. 7: Περιγράφεται σχηματικά η λειτουργία των GANs.

2.9.2 Αρχιτεκτονικές Παραγωγικών Ανταγωνιστικών Δικτύων

2.9.2.1 Πλήρως συνδεδεμένα Παραγωγικά Ανταγωνιστικά Δίκτυα

Οι πρώτες αρχιτεκτονικές για την δημιουργία των GAN χρησιμοποιούσαν πλήρως συνδεδεμένα νευρωνικά δίκτυα (fully connected GANs) και για τα δυο νευρωνικά δίκτυα (generator, discriminator). Σε αυτή την αρχιτεκτονική κάθε νευρώνας των κρυφών επιπέδων συνδέεται με την έξοδο όλων των νευρώνων του προηγούμενου επιπέδου [44]. Ωστόσο αυτή η αρχιτεκτονική εφαρμόστηκε για την δημιουργία εικόνων σχετικά απλών όπως για τη δημιουργία χειρόγραφων ψηφίων (MNIST), φυσικών εικόνων (CIFAR-10) και εικόνων δεδομένων προσώπου του Τορόντο (TFD).

2.9.2.2 Συνελκτικά Παραγωγικά Ανταγωνιστικά Δίκτυα

Τα πρώτα πειράματα για τα συνελκτικά παραγωγικά ανταγωνιστικά δίκτυα ή αλλιώς convolutional GANs έγιναν πάνω στην βάση δεδομένων των φυσικών εικόνων (CIFAR-10). Η διαδικασία αυτή έδειξε ότι ήταν πιο δύσκολο να εκπαιδευτούν ο δημιουργός και ο διευκρινιστής μέσω των συνελκτικών δικτύων.

2.9.2.3 Υπό όρους Παραγωγικά Ανταγωνιστικά Δίκτυα

Τα παραγωγικά ανταγωνιστικά δίκτυα μπορούν να επεκταθούν σε “υπό όρους” δίκτυα (conditional GANs) εάν και ο δημιουργός (generator) και ο διευκρινιστής (discriminator) εξαρτώνται από κάποιες επιπλέον βοηθητικές πληροφορίες, όπως για παράδειγμα ετικέτες (lables). Αυτό επιτυγχάνεται τροφοδοτώντας και τα δύο αυτά δίκτυα του generator και discriminator με ένα επιπλέον στρώμα εισόδου. Το πλεονέκτημα αυτής της τεχνικής είναι ότι παρέχει καλύτερες αναπαραστάσεις για δημιουργίες πολυτροπικών δεδομένων (multimodal data) [46].

2.9.2.4 GAN με μοντέλα συμπερασμάτων

Αρχικά τα GAN δεν μπορούσαν να χαρτογραφήσουν την δοσμένη είσοδο κάτι το οποίο χρειαζόταν. Έχουν προταθεί διάφορες τεχνικές για την δημιουργία των προ-εκπαιδευμένων GANs (pretrained GANs), ωστόσο τα adversarially learned inference (ALI) και τα Bidirectional GANs (BiGANs) προτείνουν ένα αποτελεσματικό δίκτυο αποτελεσμάτων. Σύμφωνα με αυτό ο διευκρινιστής (discriminator) ελέγχει κοινά ζεύγη από τα πραγματικά δεδομένα και τα νεοσύστατα δεδομένα. Συγκεκριμένα ο δημιουργός αποτελείται από δύο δίκτυα, τον ‘κωδικοποιητή’ και τον ‘αποκωδικοποιητή’. Και τα δύο αυτά δίκτυα εκπαιδεύονται μαζί για να ξεγελάσουν τον discriminator. Έπειτα ο τελευταίος λαμβάνει ζεύγη διανυσμάτων και οφείλει να καθορίσει ποιο ζεύγος προέρχεται από πραγματικό δείγμα εικόνας μαζί με την κωδικοποίηση του και ποιο από πλαστό. Παρατηρείται ότι τα νέα δείγματα που δημιουργούνται με την χρήση του ALI και του BiGAN δεν είναι καλή, αλλά βελτιώνεται με ένα επιπλέον κόστος.

2.9.2.5 Adversarial Autoencoders

Οι autoencoders είναι δίκτυα που αποτελούνται από έναν κωδικοποιητή (encoder) και έναν αποκωδικοποιητή (decoder) που εκπαιδεύεται να χαρτογραφεί δεδομένα σε μη πραγματικές αναπαραστάσεις. Ειδικότερα τα δίκτυα αυτά μέσω του κωδικοποιητή μαθαίνουν ντετερμινιστικά να χαρτογραφούν από πραγματικά σε μη πραγματικά δεδομένα και μέσω του αποκωδικοποιητή το ανάποδο. Φυσικά σκοπός και των δύο αυτών δικτύων είναι η δημιουργία μια νέας εικόνας που θα πλησιάζει όσο το δυνατόν το πρωτότυπο. Αξίζει να αναφερθεί πως οι autoencoders χαρτογραφούν μη γραμμικά και στις δυο διαστάσεις. Επιπλέον έχουν μεγάλη ευελιξία όταν υλοποιούνται με βαθιά νευρωνικά δίκτυα (deep networks) ενώ η εκπαίδευσή τους μπορεί να γίνει χωρίς επίβλεψη (unsupervised).

2.10 Υπερπροσαρμογή

Η υπερπροσαρμογή (overfitting) είναι ένα πρόβλημα που αντιμετωπίζει η μηχανική μάθηση. Ειδικότερα όταν το μοντέλο υφίσταται υπερπροσαρμογή πρακτικά σημαίνει ότι το δίκτυο δεν μαθαίνει χρήσιμες πληροφορίες από το σύνολο δεδομένων που του παρέχονται για να εκπαιδευτεί αλλά “απομνημονεύει” τις πληροφορίες ή εστιάζει σε ανούσιες λεπτομέρειες. Για να εντοπιστεί το πρόβλημα της υπερπροσαρμογής συνηθίζεται να δημιουργείται και ένα δεύτερο σύνολο, εκτός από εκείνο της εκπαίδευσης, το σύνολο της επικύρωσης (validation dataset). Ουσιαστικά ένα μικρό ποσοστό του συνόλου εκπαίδευσης αποκόπτεται από το σύνολο αυτό και δημιουργεί το σύνολο επικύρωσης. Το overfitting παρατηρείται όταν η μέση συνάρτηση σφάλματος επιβεβαίωσης είναι αρκετά υψηλότερη από την απώλεια εκπαίδευσης ή αντιστρόφως εάν η ακρίβεια της εκπαίδευσης είναι αρκετά καλύτερη από εκείνη της επιβεβαίωσης. Για να αποφευχθεί το πρόβλημα της υπερπροσαρμογής μια λύση είναι η διακοπή της εκπαίδευσής του όταν υπάρχει μεγάλη διαφορά στις απώλειες ή την απόδοση ανάμεσα στα δύο σύνολα (training και validation dataset) [47].

2.11 Αξιολόγηση των Παραγωγικών Ανταγωνιστικών Δικτύων

Τα Generative adversarial networks έχουν γίνει ένας από τους πιο δημοφιλείς τρόπους για την δημιουργία εικόνων στις μέρες μας. Είναι ευρέως διαδεδομένα για τις πολλών ειδών παραγόμενες εικόνες που δημιουργούν. Μπορούν όμως αυτές οι παραγόμενες εικόνες να συγκριθούν με πραγματικές εικόνες; Μετά την δημιουργία τέτοιων εικόνων προκύπτει ένα βασικό ερώτημα το οποίο είναι αν οι εικόνες αυτές πληρούν ορισμένα κριτήρια. Εάν δηλαδή το GAN είναι ικανό να δημιουργήσει τις εικόνες που αναμένουμε ότι θα πρέπει να δημιουργήσει. Για παράδειγμα εάν το GAN έχει εκπαιδευτεί για να δημιουργεί εικόνες από γάτες, οι εικόνες που θα παράγει θα πρέπει να μοιάζουν με γάτες. Όπως είναι αντιληπτό η αξιολόγηση και η σύγκριση δύο η περισσότερων GAN μεταξύ τους είναι μία πρόκληση. Φυσικά πολλές φορές η αξιολόγηση των νευρωνικών αυτών μοντέλων γίνεται μόνο μέσω υποκειμενικής οπτικής αξιολόγησης των παραγόμενων εικόνων από κάποιο ανθρώπινο μάτι. Φυσικά ένας πιο αντικειμενικός τρόπος είναι πάντα πιο αξιόπιστος και για αυτό παρακάτω παρουσιάζονται κάποιες ποσοτικές μετρικές για την αξιολόγηση των GANs. Φυσικά η αξιολόγηση των GAN γίνεται μέσω των παραγόμενων εικόνων που δημιουργούν.

2.11.1 GAN- train Scores

Για να αξιολογήσουμε μέσω των GAN - train Scores τις παραγόμενες εικόνες του GAN και κατ' επέκταση το ίδιο το GAN εκπαιδευόμενα ένα δίκτυο ταξινόμησης με εικόνες που έχουν δημιουργηθεί από το GAN που θέλουμε να αξιολογήσουμε. Στην συνέχεια αφού το μοντέλο εκπαιδευτεί με τις παραγόμενες εικόνες, εισάγονται σε αυτό πραγματικές εικόνες προκειμένου να αξιολογηθεί η απόδοσή του. Αξίζει να σημειωθεί ότι το δοκιμαστικό σύνολο των πραγματικών εικόνων που θα χρησιμοποιήσει το δίκτυο πρέπει να είναι διαφορετικό από εκείνο που χρησιμοποιήθηκε από το GAN για να εκπαιδευτεί. Συγκεκριμένα το GAN - train είναι η ακρίβεια ενός ταξινομητή που έχει εκπαιδευτεί και αξιολογηθεί στα αντίστοιχα σύνολα. Με αυτό τον τρόπο μετρείται η διαφορά μεταξύ της κατανομής της μάθησης, δηλαδή των παραγόμενων εικόνων, και των πραγματικών εικόνων του δοκιμαστικού συνόλου. Μπορούμε να συμπεράνουμε ότι οι παραγόμενες εικόνες μοιάζουν με τις πραγματικές, δηλαδή ότι το GAN έχει καλή απόδοση, εάν το μοντέλο ταξινόμησης μπορέσει να ταξινομήσει σωστά τις πραγματικές εικόνες του δοκιμαστικού συνόλου. Όσο χαμηλότερη είναι η ακρίβεια τόσο χειρότερες εικόνες έχει παράγει το GAN. Οι λόγοι που η ακρίβεια μπορεί να είναι χαμηλότερη είναι πολλοί και κάποιοι από αυτούς είναι η ποικιλομορφία των συνόλων εκπαίδευσης και αξιολόγησης ή το γεγονός τα παραγόμενα δείγματα του GAN δεν είναι αρκετά ρεαλιστικά ώστε να εκπαιδεύσουν επαρκώς τον ταξινομητή. Επίσης άλλος ένας παράγοντας που μπορεί να οφείλεται η χαμηλή απόδοση του GAN- train είναι ότι το GAN μπορεί να έχει ανακατέψει διαφορετικές κλάσεις μεταξύ τους και αυτό να μπερδεύει τον ταξινομητή να κάνει σωστή κατηγοριοποίηση. Τέλος αξίζει να αναφερθεί ότι οι αποτυχίες του GAN είναι δύσκολο να διαγνωστούν. Επίσης όταν η ακρίβεια πάρει μεγάλες τιμές συμπεραίνεται ότι οι εικόνες είναι υψηλής ποιότητας και διαφέρουν μεταξύ τους. Φυσικά η ποικιλομορφία ποικίλλει σύμφωνα με τον συνολικό αριθμό των παραγόμενων εικόνων του dataset. Όσο μεγαλύτερο το σύνολο τόσο περισσότερες είναι οι πιθανότητες να ποικίλουν και οι εικόνες μεταξύ τους [48].

2.11.2 GAN- test Scores

Άλλη μια μετρική για την αξιολόγηση ενός GAN είναι το GAN- test. Το GAN- test μετρά την ακρίβεια ενός δικτύου ταξινόμησης που έχει εκπαιδευτεί σε πραγματικές εικόνες και αξιολογείται από τις παραγόμενες εικόνες του GAN. Αυτή η μετρική ουσιαστικά υποδηλώνει ότι οι παραγόμενες εικόνες αποτελούν μια ρεαλιστική προσέγγιση της κατανομής των πραγματικών εικόνων. Συγκεκριμένα η μετρική GAN – test είναι η ακρίβεια ενός ταξινομητή που έχει εκπαιδευτεί και αξιολογηθεί στα αντίστοιχα σύνολα. Με αυτό τον τρόπο μετριέται η διαφορά μεταξύ της κατανομής της μάθησης, δηλαδή των πραγματικών εικόνων, και των παραγόμενων εικόνων του δοκιμαστικού συνόλου. Σκοπός είναι η ακρίβεια του μοντέλου, όταν εκείνο αξιολογήθηκε με τις παραγόμενες εικόνες, να είναι όσο το δυνατόν πιο κοντά στην ακρίβεια του μοντέλου όταν εκείνο αξιολογήθηκε με το validation dataset. Το validation dataset περιλαμβάνει πραγματικές εικόνες, διαφορετικές όμως από εκείνες που χρησιμοποιήθηκαν στην εκπαίδευση. Μπορούμε να συμπεράνουμε ότι οι παραγόμενες εικόνες μοιάζουν με τις πραγματικές, δηλαδή ότι το GAN έχει καλή απόδοση, εάν το μοντέλο ταξινόμησης μπορέσει να ταξινομήσει σωστά τις παραγόμενες εικόνες του δοκιμαστικού συνόλου και η ακρίβεια του ήταν πολύ κοντά στην ακρίβεια του validation dataset. Όσο χαμηλότερη είναι η ακρίβεια τόσο χειρότερες εικόνες έχει παράγει το GAN. Στην περίπτωση που η ακρίβεια του μοντέλου ξεπεράσει την ακρίβεια που έχει το μοντέλο όταν αξιολογηθεί με το validation dataset, σημαίνει ότι το GAN απλά απομνημονεύει και αντιγράφει της πραγματικές εικόνες εκπαίδευσης, δημιουργώντας αντίγραφα αυτών. Αξίζει να ειπωθεί ότι αυτή η μετρική δεν καταγράφει την ποικιλομορφία των εικόνων, αφού ένα μοντέλο που απομνημόνευσε μια εικόνα εκπαίδευσης η οποία έχει παρόμοια κατανομή με τις παραγόμενες θα έχει πολύ καλή απόδοση [48].

2.11.3 Μήτρα σύγχυσης

Η μήτρα σύγχυσης (confusion matrix) είναι μία πολύ δημοφιλής μετρική για προβλήματα ταξινόμησης. Συγκεκριμένα είναι ένας πίνακας που χρησιμοποιείται για να αναπαραστήσει την απόδοση ενός αλγόριθμου ταξινόμησης. Καταφέρει να οπτικοποιήσει και να συγκεντρώσει πόσο καλά έλυσε ο αλγόριθμος το εκάστοτε πρόβλημα. Με λίγα λόγια ποσοτικοποιήσει τον αριθμό των σωστά προβλεπόμενων αποτελεσμάτων αλλά και εκείνων που δεν κατάφερε ο ταξινομητής να ταξινομήσει σωστά. Ας υποθέσουμε ότι έχουμε ένα πρόβλημα ταξινόμησης δύο κλάσεων. Αρχικά ονομάζει την μία κλάση ως “Θετική” (Positive) και την δεύτερη ως “Αρνητική” (Negative). Έπειτα ο confusion matrix είναι ένας πίνακας 2 επί 2. Την μια θέση του πίνακα την καταλαμβάνει η τιμή του “Αληθώς θετικό” (True Positive -TP). Η τιμή αυτή αναπαριστά το νούμερο των “Θετικών” (Positive) που αναγνωρίστηκαν επιτυχώς ως “Θετικά”. Την δεύτερη θέση τιμή του πίνακα καταλαμβάνει η τιμή του “Αληθώς αρνητικό” (True Negative - TN). Η τιμή αυτή αναπαριστά το νούμερο των “Αρνητικών” (Negative) που αναγνωρίστηκαν σωστά από τον ταξινομητή ως “Αρνητικά”. Την τρίτη θέση του πίνακα παίρνει η τιμή του “Ψευδώς θετικό” (False Positive - FP). Η τιμή αυτή δείχνει τον αριθμό των “Αρνητικών” που λανθασμένα αναγνωρίστηκαν ως “Θετικά” από τον ταξινομητή. Τέλος η τελευταία θέση του confusion matrix ανήκει στην τιμή του “Ψευδώς αρνητικό” (False Negative -FN). Η τιμή αυτή φανερώνει τον αριθμό των “Θετικών” που λανθασμένα αναγνωρίστηκαν ως “Αρνητικά”. Έχοντας υπολογίσει τις τιμές TP, TN, FP και FN μπορούμε στην συνέχεια υπολογίσουμε κάποιες μετρικές που θα μας δώσουν πληροφορίες για την απόδοση του μοντέλου [49] [50].

2.11.4 True Positive Rate

Το True Positive Rate (TPR) ή αλλιώς sensitivity, recall ή hit rate είναι μια μετρική που όπως προδίδει και το όνομα της φανερώνει την ευαισθησία του μοντέλου και ειδικότερα πόσο καλά μπορεί να αναγνωρίσει την κλάση των Positive. Συγκεκριμένα ορίζεται ως η αναλογία των σωστά ταξινομημένων Positive (TP) διαιρούμενη με τον συνολικό αριθμό των στοιχείων που ανήκουν στην κλάση των Positive.

Άρα

$$TPR = \frac{TP}{P} = \frac{TP}{TP + FN}$$

Η δεύτερη ισότητα ισχύει γιατί το TP , δηλαδή τα σωστά ταξινομημένα Positive, και τα FN ,δηλαδή τα λανθασμένα ταξινομημένα ως Negative τα οποία δηλαδή είναι Positive, αν αθροιστούν μας δίνουν το σύνολο των Positive στοιχείων. Εάν η μετρική πάρει τιμή ίση με το 1 σημαίνει ότι ταξινομήθηκαν όλα τα Positive σωστά, ενώ εάν πάρει τιμή 0 σημαίνει ότι όλα τα Positive ταξινομήθηκαν λάθος δηλαδή ως Negative [49] [50].

2.11.5 True Negative Rate

Το True Negative Rate (TNR) ή αλλιώς specificity ή selectivity είναι και αυτό μια μετρική αξιολόγησης της απόδοσης παρόμοια με την True Positive Rate η οποία μετράει πόσο καλά μπορεί το μοντέλο να αναγνωρίσει την κλάση των Negative. Συγκεκριμένα ορίζεται ως η αναλογία των σωστά ταξινομημένων Negative (TN) διαιρούμενη με τον συνολικό αριθμό των στοιχείων που ανήκουν στην κλάση των Negative.

Άρα

$$TNR = \frac{TN}{N} = \frac{TN}{TN + FP}$$

Η δεύτερη ισότητα ισχύει γιατί το TN , δηλαδή τα σωστά ταξινομημένα Negative και τα FP ,δηλαδή τα λανθασμένα ταξινομημένα ως Positive τα οποία δηλαδή είναι Negative, αν αθροιστούν μας δίνουν το σύνολο των Negative στοιχείων. Εάν η μετρική πάρει τιμή ίση με το 1 σημαίνει ότι ταξινομήθηκαν όλα τα Negative σωστά, ενώ εάν πάρει τιμή 0 σημαίνει ότι όλα τα Negative ταξινομήθηκαν λάθος δηλαδή ως Positive [49] [50].

2.11.6 False Negative Rate

Το False Negative Rate (FNR) ή αλλιώς miss rate είναι αντίστροφη μετρική αξιολόγησης του True Positive Rate, αναγνωρίζει δηλαδή το ποσοστό των Positive τα οποία αναγνωρίστηκαν λανθασμένα ως Negative. Συγκεκριμένα ορίζεται ως η αναλογία των λανθασμένα ταξινομημένων Positive (FN) διαιρούμενη με τον συνολικό αριθμό των στοιχείων που ανήκουν στην κλάση των Positive.

Άρα

$$FNR = \frac{FN}{P} = \frac{FN}{FN + TP} = 1 - TPR$$

Η δεύτερη ισότητα ισχύει γιατί το TP , δηλαδή τα σωστά ταξινομημένα Positive και τα FN ,δηλαδή τα λανθασμένα ταξινομημένα ως Negative τα οποία δηλαδή είναι Positive, αν αθροιστούν μας δίνουν το σύνολο των Positive στοιχείων. Επίσης επειδή τα FNR και TPR είναι αντιστρόφως ανάλογα, το άθροισμά τους ισούται με μονάδα. Εάν η μετρική πάρει τιμή ίση με το 1 σημαίνει ότι ταξινομήθηκαν όλα τα Positive λάθος, ενώ εάν πάρει τιμή 0 σημαίνει ότι όλα τα Positive ταξινομήθηκαν σωστά δηλαδή ως Positive [49] [50].

2.11.7 False Positive Rate

Το False Positive Rate (FPR) ή αλλιώς fall-out είναι αντίστροφη μετρική αξιολόγησης του True Negative Rate, αναγνωρίζει δηλαδή το ποσοστό των Negative τα οποία αναγνωρίστηκαν λανθασμένα ως Positive Συγκεκριμένα ορίζεται ως η αναλογία των λανθασμένα ταξινομημένων Negative (FP) διαιρούμενη με τον συνολικό αριθμό των στοιχείων που ανήκουν στην κλάση των Negative.

Άρα

$$FPR = \frac{FP}{N} = \frac{FP}{FP + TN} = 1 - TNR$$

Η δεύτερη ισότητα ισχύει γιατί το TN , δηλαδή τα σωστά ταξινομημένα Negative, και τα FP ,δηλαδή τα λανθασμένα ταξινομημένα ως Positive τα οποία δηλαδή είναι Negative, αν αθροιστούν μας δίνουν το σύνολο των Negative στοιχείων. Επίσης επειδή τα FPR και TNR είναι αντιστρόφως ανάλογα, το άθροισμά τους

ισούται με μονάδα. Εάν η μετρική πάρει τιμή ίση με το 1 σημαίνει ότι ταξινομήθηκαν όλα τα Negative λάθος, ενώ εάν πάρει τιμή 0 σημαίνει ότι όλα τα Negative ταξινομήθηκαν σωστά δηλαδή ως Negative [49][50].

2.11.8 Accuracy

Η μετρική αυτή είναι η πιο κοινώς χρησιμοποιούμενη για να αξιολογηθεί η απόδοση της ταξινόμησης. Συγκεκριμένα ορίζεται ως η αναλογία των σωστά ταξινομημένων στοιχείων προς τον συνολικό αριθμό των στοιχείων. Στην περίπτωση που αναφέρουμε και παραπάνω, το accuracy υπολογίζεται ως το άθροισμα των σωστά ταξινομημένων Positive ως Positive (TP) και των σωστά ταξινομημένων Negative ως Negative (TN) δια το άθροισμα των Negative και των Positive. Η τιμή 1 δηλώνει ότι η ταξινόμηση πέτυχε πλήρως και όλα τα Positive αναγνωρίστηκαν σαν Positive και όλα τα Negative ως Negative [49] [50].

Άρα

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} = \frac{TP + TN}{P + N}$$

2.11.9 Balanced Accuracy

Η “ακρίβεια” (Accuracy) μπορεί να είναι παραπλανητική μετρική εάν τα σύνολα δεδομένων που χρησιμοποιούνται δεν είναι ισορροπημένα, όταν δηλαδή το σύνολο των Positive είναι διαφορετικό σε μέγεθος από αυτό των Negative. Για τέτοιες περιπτώσεις υπάρχει το Balanced Accuracy (BA) το οποίο όπως και το Accuracy ορίζεται ως η αναλογία των σωστά ταξινομημένων στοιχείων προς τον συνολικό αριθμό των στοιχείων. Η διαφορά του είναι ότι κανονικοποιεί τις τιμές του True Positive και του True negative με τον αριθμό των Positive και Negative στοιχείων που υπάρχουν και διαιρεί το άθροισμά τους με το 2. Ουσιαστικά προκύπτει προσθέτοντας το TPR με το TNR και διαιρώντας τα με το 2.

Άρα

$$BA = \frac{TPR + TNR}{2}$$

Στην περίπτωση που τα σύνολα δεδομένων του Positive και του Negative είναι ίσα, τα Balanced Accuracy και Accuracy θα είναι ίσα. Η τιμή 1 δηλώνει ότι η ταξινόμηση πέτυχε πλήρως και όλα τα Positive αναγνωρίστηκαν σαν Positive και όλα τα Negative ως Negative [49] [50].

2.11.10 Positive Predictive value

Το Positive Predictive value (PPV) ή αλλιώς precision ορίζεται η αναλογία των σωστά ταξινομημένων Positive (TP) προς το σύνολο όλων των δεδομένων που αναγνωρίστηκαν ως Positive ακόμα και αν δεν είναι. Ειδικότερα υπολογίζεται διαιρώντας το TP με το σύνολο όσων ταξινομήθηκαν ως Positive. Η μέγιστη τιμή που μπορεί να πάρει είναι το 1 που δηλώνει ότι κανένα Negative δεν αναγνωρίστηκε ως Positive και η ελάχιστη το 0 που σημαίνει ότι κανένα Positive δεν αναγνωρίστηκε σωστά ως Positive [49] [50].

Άρα

$$PPV = \frac{TP}{TP + FP}$$

2.11.11 Negative Predictive value

Το Negative Predictive value (NPV) ορίζεται η αναλογία των σωστά ταξινομημένων Negative (TN) προς το σύνολο όλων των δεδομένων που αναγνωρίστηκαν ως Negative ακόμα και αν δεν είναι. Ειδικότερα

υπολογίζεται διαιρώντας το TN με το σύνολο όσων ταξινομήθηκαν ως Negative. Η μέγιστη τιμή που μπορεί να πάρει είναι το 1 που δηλώνει ότι κανένα Positive δεν αναγνωρίστηκε ως Negative και η ελάχιστη το 0 που σημαίνει ότι κανένα Negative δεν αναγνωρίστηκε σωστά ως Negative [49] [50].

Άρα

$$NPV = \frac{TN}{TN + FN}$$

2.11.12 F1 score

Η μετρική F1 score ή αλλιώς F είναι και αυτή γνωστή για την αξιολόγηση των ταξινομήσεων. Ειδικότερα η μετρική αυτή δηλώνει την ισορροπία ανάμεσα στο Positive Predictive value και στο True Positive Rate. Ουσιαστικά είναι ο αρμονικός μέσος όρος του precision και του recall. Η υψηλότερη βαθμολογία που μπορεί να πάρει είναι το 1, φανερώνοντας τέλεια precision και recall. Επίσης η μικρότερη τιμή είναι το μηδέν που μπορεί να πάρει εάν είτε το precision είτε το recall είναι μηδέν [49] [50].

Άρα

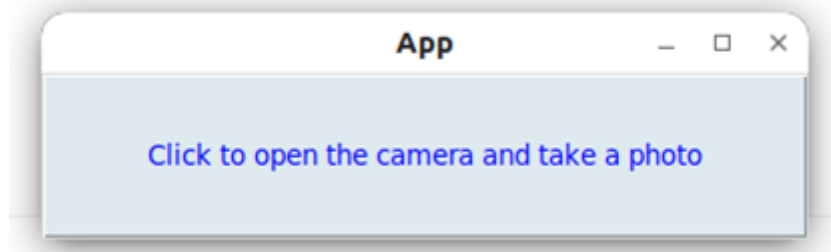
$$F1\ score = \frac{2\ PPV * TPR}{PPV + TPR} = \frac{2TP}{2TP + FP + FN}$$

Κεφάλαιο 3: Πειραματική Διαδικασία

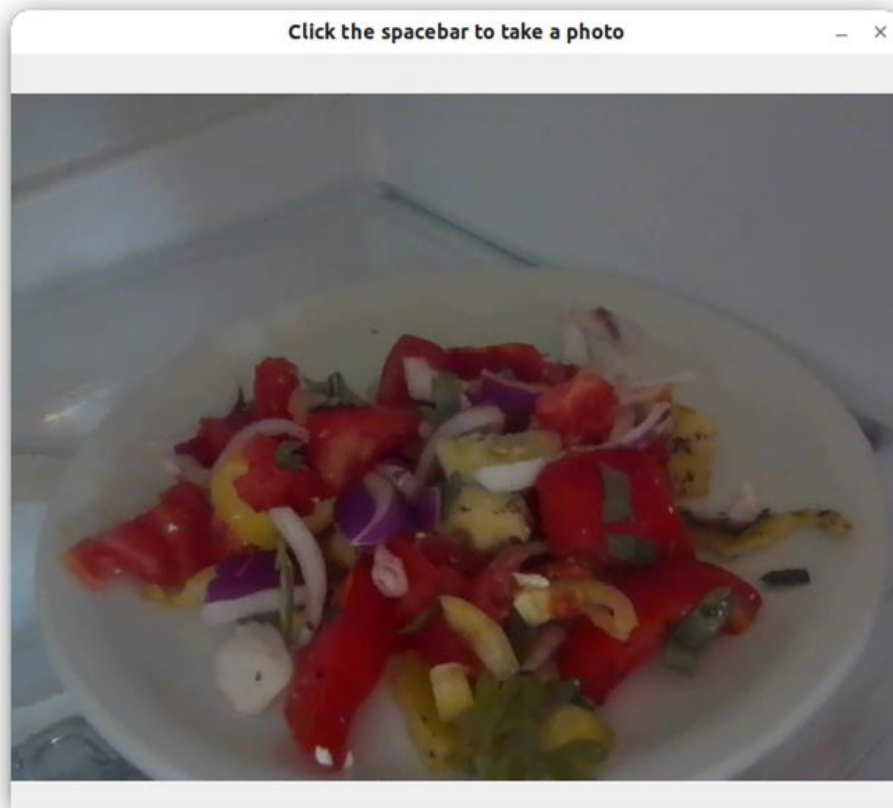
Σκοπός της παρούσας εργασίας είναι να δημιουργηθεί μια εφαρμογή με την οποία ο χρήστης θα μπορεί να πάρει μια συνθετική εικόνα στην οποία θα απεικονίζεται το αποτέλεσμα της μαγειρικής του, εάν μαγειρέψει με τα ήδη υπάρχοντα συστατικά που βρίσκονται στο ψυγείο του. Ειδικότερα ο χρήστης θα έχει την δυνατότητα πατώντας ένα κουμπί, να ενεργοποιεί μια φορητή κάμερα στο ψυγείο και έπειτα πατώντας ένα δεύτερο κουμπί να μπορεί να φωτογραφήσει το περιεχόμενο του ψυγείου του ή του ντουλαπιού του. Έπειτα η φωτογραφία αυτή θα στέλνεται στην πλατφόρμα της Google (Google Cloud Platform) και συγκεκριμένα στο Google Cloud's Vision API. Με την βοήθεια του Google Cloud Platform καταφέρνουμε να αναγνωρίσουμε και να έχουμε σε γραπτή πλέον μορφή τα αντικείμενα που υπάρχουν στην φωτογραφία που τράβηξε ο χρήστης. Μέσω λοιπόν της οπτικής αναγνώρισης που κάνει η πλατφόρμα της Google μας επιστρέφονται οι “ετικέτες” (labels) που χαρακτηρίζουν τα αντικείμενα της φωτογραφίας. Έτσι έχουμε την δυνατότητα να αναγνωρίσουμε και να ξεχωρίσουμε τα υλικά που διαθέτει ο χρήστης στο ψυγείο του και μπορούν να αξιοποιηθούν για να μαγειρέψει. Έπειτα σκοπός ήταν να εκπαιδευτεί ένα Generative Adversarial Network (GAN) με εικόνες που χαρακτηρίζονται από τις “ετικέτες” ώστε τελικώς να δημιουργήσει μια νέα παραγόμενη εικόνα η οποία θα έχει απεικονίσει το τι θα μπορούσε να μαγειρέψει ο χρήστης αν χρησιμοποιούσε τα υλικά που έχει στην διάθεση του.

3.1 Graphical User Interface

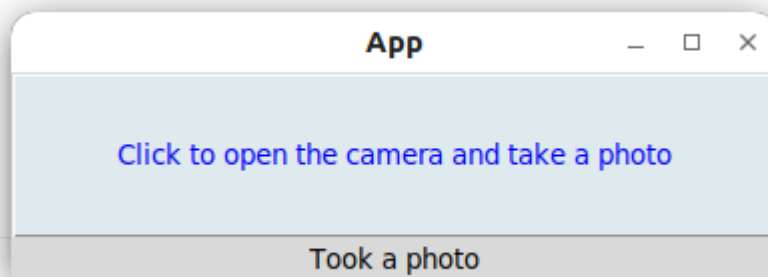
Για να υλοποιηθεί ο σκοπός της εργασίας αυτής, όπως αναφέρθηκε και παραπάνω, αρχικά δημιουργήσαμε ένα “γραφικό περιβάλλον χρήστη” (Graphical User Interface - GUI). Ειδικότερα χρησιμοποιήθηκε μια γνωστή βιβλιοθήκη της Python για την δημιουργία GUI, η “Tkinter”. Όπως φαίνεται και από την εικόνα 3.1 το GUI ουσιαστικά είναι ένα “παράθυρο” που αποτελείται από ένα κουμπί. Το κουμπί αυτό είναι υπεύθυνο για το άνοιγμα της κάμερας και επομένως με το πάτημά του από το χρήστη ανοίγει η κάμερα μέσα από ένα νέο παράθυρο. Όπως φαίνεται και στην εικόνα 3.2 εάν ο χρήστης επιθυμεί να βγάλει την φωτογραφία μπορεί να πατήσει το “spacebar” κουμπί του πληκτρολογίου. Στην περίπτωση λοιπόν που πατηθεί το spacebar και ληφθεί η φωτογραφία, ο χρήστης θα ενημερωθεί με αντίστοιχο μήνυμα ότι η φωτογραφία λήφθηκε όπως φαίνεται και από την εικόνα 3.3.



Εικόνα 3. 1: Απεικονίζεται το GUI και βλέπουμε το κουμπί που εάν πατηθεί θα ανοίξει και θα ενεργοποιηθεί η κάμερα του υπολογιστή.



Εικόνα 3. 2: Απεικονίζεται το παράθυρο με την κάμερα του υπολογιστή που εάν πατηθεί το "spacebar" θα ληφθεί η φωτογραφία.



Εικόνα 3. 3: Απεικονίζεται το GUI και βλέπουμε το κουμπί που εάν πατηθεί θα ανοίξει και θα ενεργοποιηθεί η κάμερα του υπολογιστή αλλά και το σχετικό μήνυμα που ενημερώνει το χρήστη ότι η φωτογραφία λήφθηκε.

3.2 Google Cloud's Vision API

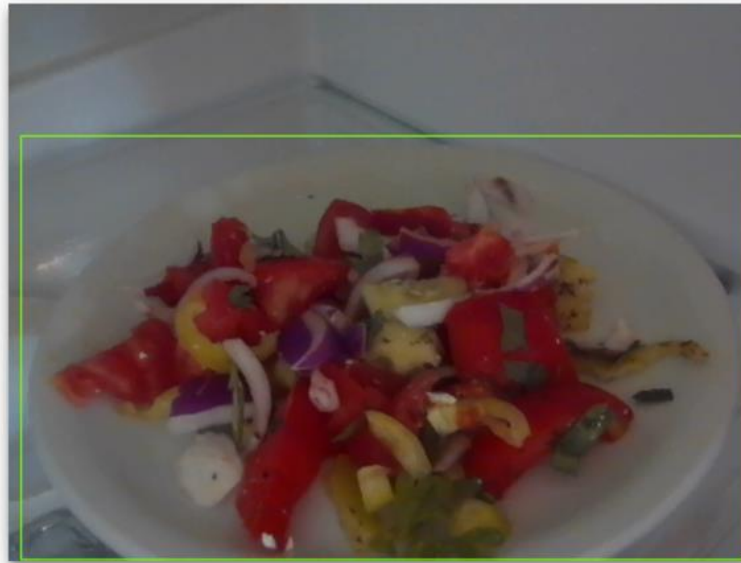
Όπως ήδη προαναφέρθηκε το Google Cloud's Vision API έχει την δυνατότητα να αναγνωρίζει τα αντικείμενα που απεικονίζονται σε οποιαδήποτε εικόνα. Αξιοποιώντας αυτήν την λειτουργία, εισάγουμε την φωτογραφία που τραβήχτηκε από το ψυγείο ή το ράφι μας στο Vision API, με σκοπό να λάβουμε σε μορφή "ετικετών" τα αντικείμενα της εικόνας. Συγκεκριμένα τα τρόφιμα που υπάρχουν στο ψυγείο θα αποτυπωθούν στην φωτογραφία και με την βοήθεια του Google Cloud's Vision API θα μας επιστραφούν σε μορφή κειμένου-ετικετών (labels) ώστε να μπορέσουμε έτσι να τα χρησιμοποιήσουμε στο GAN και να συνθέσουμε την "νέα" εικόνα, που είναι και το ζητούμενο της εργασίας. Φυσικά για να καταφέρουμε να χρησιμοποιήσουμε την

πλατφόρμα της Google χρειάζεται να δημιουργήσουμε έναν λογαριασμό ώστε να λάβουμε ένα μοναδικό “κλειδί” που χωρίς αυτό είναι αδύνατη η χρήση του Vision API. Έτσι δημιουργώντας τον νέο λογαριασμό και αποθηκεύοντας το απαραίτητο “κλειδί” για την χρήση του, μπορούμε να εκτελέσουμε οποιαδήποτε οπτική αναγνώριση θέλουμε.

Λαμβάνοντας λοιπόν την παραπάνω φωτογραφία από το GUI, την αποστέλλουμε κατευθείαν στο Google Cloud’s Vision API και μας επιστρέφει ορισμένες ετικέτες (labels) και ένα ποσοστό για κάθε ετικέτα. Αρχικά όπως φαίνεται και στην εικόνα 3.4 το περιεχόμενο που αναγνωρίζει το Vision API περικλείεται μέσα στο πράσινο περίγραμμα που όπως είναι φανερό είναι και το περιεχόμενο που περιέχει την πληροφορία που θέλουμε να αναλυθεί. Συγκεκριμένα επιστρέφεται ένα JSON αρχείο που μέσα από αυτό μας ενδιαφέρουν τα πεδία “description” και “score”. Οι ετικέτες αυτές είναι το αποτέλεσμα της ανάλυσης που έκανε στην εικόνα και ουσιαστικά περιγράφουν την πιθανότητα με την οποία το Vision API εκτιμάει ότι η κάθε ετικέτα εντοπίζεται στην εικόνα. Αυτή η πιθανότητα αποτυπώνεται από το “score”, δηλαδή το ποσοστό που δηλώνει κατά πόσο η εικόνα ή κάποιο αντικείμενο της εικόνας μπορεί να περιγράφεται από την ετικέτα αυτή σύμφωνα πάντα με το Vision API όπως φαίνεται και στην εικόνα 3.5. Δηλαδή το “score” αυτό δηλώνει με πόση “βεβαιότητα” το API αυτό θεωρεί σωστή την πρόβλεψή του. Επιπροσθέτως αξίζει να αναφερθεί ότι επιστρέφονται μόνο οι ετικέτες οι οποίες συγκεντρώνουν βαθμολογία (score) μεγαλύτερη ή ίση του 50%. Στην περίπτωση της εικόνας 3.2 που λήφθηκε παραπάνω η πλατφόρμα της Google μας επιστρέφει τις εξής ετικέτες:

- “Food” (Φαγητό) με ποσοστό 98%
- “Plant” (Φυτό) με ποσοστό 93%
- “Tableware” (Επιτραπέζια Σκεύη) με ποσοστό 92%
- “Fruit” (Φρούτο) με ποσοστό 92%
- “Ingredient” (Συστατικό) με ποσοστό 89%
- “Salad” (Σαλάτα) με ποσοστό 88%
- “Dishware” (Πιάτο) με ποσοστό 88%
- “Recipe” (Συνταγή) με ποσοστό 87%
- “Fruit Salad” (Σαλάτα λαχανικών) με ποσοστό 82%
- “Cuisine” (Κουζίνα) με ποσοστό 81%
- “Tomato” (Ντομάτα) με ποσοστό 81%
- “Garnish” (Γαρνιτούρα) με ποσοστό 80%
- “Vegetable” (Λαχανικά) με ποσοστό 79%
- “Dish” (Πιάτο) με ποσοστό 78%
- “Superfood” (Υπερτροφία) με ποσοστό 77%
- “Produce” (Καρπός) με ποσοστό 76%
- “Plum Tomato” (Ντομάτα) με ποσοστό 76%
- “Natural Foods” (Φυσικές τροφές) με ποσοστό 75%
- “Platter” (Πιατέλα) με ποσοστό 72%
- “Bush Tomato” (Ντομάτα) με ποσοστό 71%
- “Sweetness” (Γλυκά) με ποσοστό 63%
- “Carrot” (Καρότα) με ποσοστό 62%
- “Vegan Nutrition” (Χορτοφαγική Διατροφή) με ποσοστό 62%
- “Plate” (Πιάτο) με ποσοστό 62%
- “Mixture” (Μείγμα) με ποσοστό 61%
- “Leaf Vegetable” (Φύλλο λαχανικών) με ποσοστό 60%
- “Side Dish” (Συνοδευτικό) με ποσοστό 59%
- “A La Carte Food” (A La Carte Φαγητό) με ποσοστό 58%
- “Strawberry” (Φράουλα) με ποσοστό 58%
- “Breakfast” (Πρωινό) με ποσοστό 57%
- “Fruit Cup” (Φλιτζάνι φρούτων) με ποσοστό 56%
- “Tursu” (Τουρσί) με ποσοστό 55%
- “Meal” (Γεύμα) με ποσοστό 54%
- “Lunch” (Μεσημεριανό) με ποσοστό 53%
- “Solanum” με ποσοστό 51%
- “Comfort Food” με ποσοστό 51%

- “Cherry Tomatoes” (Ντοματίνια) με ποσοστό 50%



Εικόνα 3. 4: Απεικονίζεται η οπτική απεικόνιση που πραγματοποιήθηκε από το Google Cloud’s Vision API στην εικόνα που λήφθηκε από το GUI. Φαίνεται ότι το Vision API αναγνώρισε ότι το περιεχόμενο της εικόνας είναι σχετικό με φαγητό.

Objects **Labels** Logos Properties Safe Search

Food	98%
Plant	93%
Tableware	92%
Fruit	92%
Ingredient	89%
Salad	88%
Dishware	88%

Εικόνα 3. 5: Απεικονίζεται η οπτική απεικόνιση που πραγματοποιήθηκε από το Google Cloud’s Vision API στην εικόνα που λήφθηκε από το GUI. Παρουσιάζονται οι ετικέτες (labels) που επιστρέφονται από το Vision API σε συνδυασμό με την πιθανότητα που δίνει το API να αναγνώρισε σωστά την αντίστοιχη ετικέτα μέσα από την εικόνα. Φαίνονται μόνο τα πρώτα labels δηλαδή εκείνα με την μεγαλύτερη βαθμολογία (score).

Παρατηρούμε λοιπόν ότι το Google Cloud's Vision API επέστρεψε αρκετές ετικέτες εκ των οποίων κάποιες είναι σχετικές με το περιεχόμενο αυτής, όπως οι

- “Food”
- “Tableware”
- “Ingredient”
- “Salad”
- “Dishware”
- “Tomato”
- “Vegetable”
- “Dish”
- “Produce”
- “Plum Tomato”
- “Natural Foods”
- “Bush Tomato”
- “Vegan Nutrition”
- “Plate”
- “Mixture”
- “Leaf Vegetable”
- “Side Dish”
- “Meal”
- “Lunch”
- “Cherry Tomatoes”.

Από την άλλη πλευρά οι ετικέτες οι οποίες είναι άσχετες με το περιεχόμενο της εικόνας και ειδικότερα δεν υπάρχει κάποιο αντικείμενο στην εικόνα που να αντιστοιχεί με κάποια από τις παραπάνω ετικέτες είναι οι εξής:

- “Plant”
- “Fruit”
- “Recipe”
- “Fruit Salad”
- “Cuisine”
- “Garnish”
- “Superfood”
- “Platter”
- “Sweetness”
- “Carrot”
- “A La Carte Food”
- “Strawberry”
- “Breakfast”
- “Fruit Cup”
- “Tursu”
- “Solanum”
- “Comfort Food”

Ειδικότερα η ετικέτα “Plant” παρότι δεν περιγράφει κανένα αντικείμενο της εικόνας συγκέντρωσε σχετικά υψηλό ποσοστό (93%). Επίσης από τις ετικέτες που θεωρούνται ότι ταιριάζουν στην εικόνα δεν μας είναι όλες χρήσιμες. Σκοπός μας είναι να εντοπίσουμε τα υλικά-τρόφιμα τα οποία υπάρχουν στην φωτογραφία που τραβήχτηκε ώστε να τα χρησιμοποιήσουμε στην συνέχεια στην σύνθεση της νέας εικόνας και όχι labels όπως “Plate”, δηλαδή “πιάτο”. Όπως είναι λοιπόν λογικό οι ετικέτες “Food”, “Tableware”, “Ingredient”, “Dishware”, “Dish”, “Plate”, “Side Dish”, “Meal”, παρόλο που σωστά επιστρέφονται από το Vision API και μάλιστα με τα υψηλά ποσοστά κάποια από αυτά (93%, 92%, 89%, 88%, 78%, 62%, 59%, 54% αντίστοιχα), αφού στην φωτογραφία υπάρχει φαγητό, συστατικό, πιάτο, σκεύος κλπ., δεν είναι αυτό που αναζητάμε και

δεν μας είναι χρήσιμες οι ετικέτες αυτές στα επόμενα βήματα της εργασίας αυτής. Ωστόσο παρατηρούμε ότι η αναγνώριση που έχει κάνει το Google Cloud's Vision API είναι αρκετά καλή αφού μας έχει επιστρέψει την ετικέτα “Salad” με 88% που είναι αρκετά υψηλό ποσοστό και περιγράφει ακριβώς την εικόνα. Επίσης οι ετικέτες “Tomato” με 81% και “Vegetable” με 79% βαθμολογία περιγράφουν εξίσου καλά την εικόνα και αντικατοπτρίζεται αυτό και από την υψηλά score που λαμβάνουν.

Συμπεραίνουμε λοιπόν ότι το Google Cloud's Vision API έχει την δυνατότητα να αναγνωρίζει το περιεχόμενο της εκάστοτε εικόνας, επιστρέφοντας σε μορφή κειμένου-ετικετών τα αντικείμενα που “βλέπει”. Ωστόσο όπως φάνηκε και από το παραπάνω παράδειγμα δεν είναι απαραίτητο ότι θα είναι αλάνθαστο και ότι όλες οι ετικέτες θα είναι σωστές. Αυτό εξαρτάται φυσικά και από την λήψη που τράβηξε ο χρήστης. Όπως είναι λογικό όταν η εικόνα περιέχει πολλά άσχετα αντικείμενα πέρα των τροφίμων που θέλουμε να αναγνωρίσει, η πλατφόρμα θα επιστρέψει περισσότερες αχρείαστες ετικέτες. Επίσης είναι πολύ πιθανό όπως φάνηκε και παραπάνω να επιστραφούν ετικέτες σχετικές με την εικόνα αλλά που δεν μπορούν να φανούν χρήσιμες στην σύνθεση νέας εικόνας. Στο παράδειγμά μας φερ' ειπείν υπάρχει το πιάτο στην εικόνα πάνω στο οποίο είναι τοποθετημένα τα λαχανικά. Αν και είναι απαραίτητο το πιάτο, οφείλεται για τις ετικέτες “Dish”, “Dishware”, “Side Dish”, “Tableware”, “Plate” και “Platter”. Τέλος όσο περισσότερα τα αντικείμενα που υπάρχουν σε μία εικόνα τόσες οι πιθανότητες να “μπερδευτεί” και να δώσει περισσότερες σωστές ή λανθασμένες προβλέψεις δηλαδή labels. Για όλους τους παραπάνω λόγους χρειάζεται να γίνει μια επιλογή ετικετών όταν ληφθούν αυτές από το Vision API.

3.3 Επιλογή ετικετών

Όσο μεγαλύτερο εύρος ετικετών έχουμε τόσο μεγαλύτερη είναι η πολυπλοκότητα του να επιλέξουμε τις κατάλληλες και πιο χρήσιμες για εμάς ετικέτες. Στην παρούσα εργασία η επιλογή των ετικετών γίνεται εύκολα. Συγκεκριμένα αναζητούμε μόνο τις ετικέτες “Bread”, “Pasta”, “Vegetables”, “Meat”, “Rice” και “Potatoes” ή φράσεις που να περιέχουν κάποια από αυτές τις λέξεις όπως για παράδειγμα “Leaf Vegetable”. Βρίσκοντας λοιπόν τις ετικέτες αυτές με έναν πολύ απλό αλγόριθμο, γνωρίζουμε ποιες από τις 6 κατηγορίες απεικονίζονται στην φωτογραφία που λήφθηκε από τον χρήστη. Στην παραπάνω φωτογραφία λοιπόν (εικόνα 3.2) εφαρμόζοντας τον αλγόριθμο αυτό παίρνουμε ότι οι ετικέτες που μας αφορούν είναι οι “Vegetable” με 79% και η “Leaf Vegetable” με 60% αντίστοιχα. Έτσι καταλήγουμε στο συμπέρασμα ότι στην φωτογραφία αυτή και κατ' επέκταση στο ψυγείο μας υπάρχει μία από τις έξι αυτές κατηγορίες και ειδικότερα υπάρχουν λαχανικά (Vegetables). Η πρόβλεψη αυτή του Google Cloud's Vision API μαζί με τον αλγόριθμο για τον διαχωρισμό των ετικετών φαίνεται ότι είναι σωστή. Είναι πασιφανές από το ανθρώπινο μάτι ότι η φωτογραφία που λήφθηκε και αντικατοπτρίζει το ράφι του ψυγείου μας περιλαμβάνει μόνο λαχανικά και όχι ψωμί, κρέας, ρύζι, μακαρόνια ή πατάτες. Φυσικά πάντα υπάρχει η πιθανότητα λάθους στην παραπάνω διαδικασία. Έτσι με τον αλγόριθμο αυτό καταφέρνουμε με εύκολο και απλό τρόπο να ξεχωρίσουμε από τις ετικέτες του Google Cloud's Vision API εκείνες που είναι υλικά-συστατικά από εκείνες που δεν είναι και ειδικότερα να κρατήσουμε εκείνες που αφορούν μια από τις 6 κατηγορίες φαγητών που μας ενδιαφέρουν.

3.4 Δημιουργία εικόνας με την βοήθεια Παραγωγικών ανταγωνιστικών δικτύων

Έχοντας λοιπόν τις ετικέτες που φανερώνουν τα τρόφιμα της φωτογραφίας σκοπός μας είναι να δημιουργήσουμε με την βοήθεια κάποιου Generative Adversarial Network (GAN) μια συνθετική εικόνα που θα απεικονίζει το τι θα μπορούσε να μαγειρέψει ο χρήστης με τα υπάρχοντα υλικά που διαθέτει. Η ιδέα είναι να εκπαιδευσουμε ένα GAN με εικόνες ώστε να δημιουργεί νέες εικόνες. Δηλαδή σκοπός είναι το GAN αυτό να παίρνει σαν είσοδο εικόνες τροφίμων, να εκπαιδεύεται σε αυτές και έπειτα να δημιουργεί μια νέα εικόνα με βάση τις εικόνες που εκπαιδεύτηκε.

3.4.1 Επιλογή συνόλου δεδομένων για την εκπαίδευση του παραγωγικού ανταγωνιστικού δικτύου

Η κατηγορία των GAN που θέλουμε να χρησιμοποιήσουμε είναι εκείνη που τα GAN εκπαιδεύονται με εικόνες. Δηλαδή τα GAN αυτά χρειάζονται ένα σύνολο δεδομένων, ένα σύνολο εικόνων ώστε να καταφέρουν να τις “μελετήσουν”, δηλαδή να εκπαιδευτούν πάνω σε αυτές. Πρώτος λοιπόν στόχος είναι να ορίσουμε αυτό το σύνολο δεδομένων με το οποίο θα εκπαιδευτεί το GAN μας. Στο σημείο αυτό ερχόμαστε να ενώσουμε όσα ήδη έχουν εφαρμοστεί παραπάνω. Η εικόνα που επιζητούμε να δημιουργήσει το GAN μας, δεν θέλουμε να είναι μια οποιαδήποτε εικόνα αλλά μια εικόνα που θα απεικονίζει το αποτέλεσμα της μαγειρικής με τα υλικά που διαθέτουμε. Γι’ αυτό πρέπει και να το εκπαιδεύσουμε με το κατάλληλο κάθε φορά σύνολο δεδομένων. Με την βοήθεια του Vision API μπόρεσαμε να πάρουμε με την μορφή ετικετών τα αντικείμενα-τρόφιμα που περιέχει το ψυγείο ή το ράφι μας. Έπειτα ξεχωρίσαμε τις ετικέτες και κρατήσαμε μόνο όσες είναι “Bread”, “Pasta”, “Vegetables”, “Meat”, “Rice” ή “Potatoes”. Έτσι γνωρίζοντας αν κάποιο συστατικό που υπάρχει στο ψυγείο μας ανήκει σε κάποια ή κάποιες από αυτές τις έξι κατηγορίες μπορούμε να ορίσουμε το κατάλληλο σύνολο δεδομένων. Ειδικότερα ανάλογα με το ποια από αυτά τα έξι labels έχουν αναγνωριστεί στην εκάστοτε εικόνα, το GAN δέχεται σαν είσοδο 200 εικόνες από κάθε κατηγορία. Για παράδειγμα στο παράδειγμα που εξετάζεται παραπάνω, η εικόνα ανήκει μόνο στην κατηγορία των λαχανικών επομένως το GAN θα λάβει σαν είσοδο μόνο 200 πραγματικές εικόνες λαχανικών ώστε να εκπαιδευτεί. Επιπλέον πρέπει να αναφερθεί ότι όλες οι εικόνες και των έξι κατηγοριών είναι 300 x 300 pixels. Παρακάτω παρουσιάζονται έξι εικόνες, όπου η κάθε μία εικόνα αποτελεί μία από τις έξι κατηγορίες εικόνων που εξετάζουμε.



Εικόνα 3. 7: Ενδεικτική εικόνα λαχανικών από το σύνολο δεδομένων που εισάγεται στο GAN ώστε να εκπαιδευτεί αν έχει χαρακτηριστεί η φωτογραφία που τραβήχτηκε από το ψυγείο μας με ετικέτα “Vegetables”.



Εικόνα 3. 6: Ενδεικτική εικόνα ρυζιού από το σύνολο δεδομένων που εισάγεται στο GAN ώστε να εκπαιδευτεί αν έχει χαρακτηριστεί η φωτογραφία που τραβήχτηκε από το ψυγείο μας με ετικέτα “Rice”.



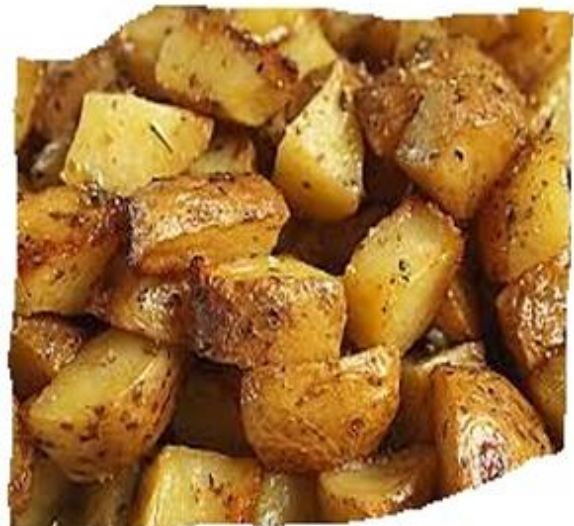
Εικόνα 3. 9: Ενδεικτική εικόνα μακαρονιών από το σύνολο δεδομένων που εισάγεται στο GAN ώστε να εκπαιδευτεί αν έχει χαρακτηριστεί η φωτογραφία που τραβήχτηκε από το ψυγείο μας με ετικέτα “Pasta”.



Εικόνα 3. 8: Ενδεικτική εικόνα ψωμιού από το σύνολο δεδομένων που εισάγεται στο GAN ώστε να εκπαιδευτεί αν έχει χαρακτηριστεί η φωτογραφία που τραβήχτηκε από το ψυγείο μας με ετικέτα “Bread”.



Εικόνα 3. 11: Ενδεικτική εικόνα κρέατος από το σύνολο δεδομένων που εισάγεται στο GAN ώστε να εκπαιδευτεί αν έχει χαρακτηριστεί η φωτογραφία που τραβήχτηκε από το ψυγείο μας με ετικέτα “Meat”.



Εικόνα 3. 10: Ενδεικτική εικόνα πατατών από το σύνολο δεδομένων που εισάγεται στο GAN ώστε να εκπαιδευτεί αν έχει χαρακτηριστεί η φωτογραφία που τραβήχτηκε από το ψυγείο μας με ετικέτα “Potatoes”.

3.4.2 Παραγωγικό ανταγωνιστικό δίκτυο: Image-Generator

Επόμενο βήμα είναι να υλοποιηθεί το GAN. Έπειτα από έρευνα βρήκαμε ένα Generative Adversarial Network το οποίο μπορεί να εκτελέσει ακριβώς αυτό που χρειαζόμαστε. Έχει δημιουργηθεί για να εκπαιδεύεται με εικόνες και ανάλογα με το σύνολο των δεδομένων που παίρνει σαν είσοδο δημιουργεί νέες παραγόμενες εικόνες. Επίσης το συγκεκριμένο GAN μπορεί να εκπαιδευτεί σε οποιοδήποτε σύνολο δεδομένων. Το GAN αυτό ονομάζεται “Image-Generator” [51][52] και αποτελείται από έναν generator και έναν discriminator όπως και όλα τα GAN. Ειδικότερα είναι ένα Deep Convolutional Generative Adversarial Network (DCGAN) δηλαδή ένα είδος των GAN που ρητά χρησιμοποιούν Convolutional και Convolutional Transpose επίπεδα στον discriminator και τον generator αντίστοιχα. Επιπλέον το GAN αυτό κάνει μη επιβλεπόμενη εκπαίδευση (Unsupervised Learning). Το συγκεκριμένο GAN διαθέτει ήδη κάποια προ-εκπαιδευμένα μοντέλα (pretrained models) τα οποία είναι διαθέσιμα για να χρησιμοποιηθούν. Γενικότερα διαθέτει κάποιες παραμέτρους εκπαίδευσης που παρόλο που έχουν ήδη ορισμένες τιμές, είναι πιθανό τροποποιώντας κάποια ή κάποιες από αυτές να βελτιωθεί η απόδοση του. Φυσικά όπως είναι λογικό οι βέλτιστες τιμές των παραμέτρων αυτών διαφέρουν ανάλογα με το πείραμα που πρόκειται να εκτελεστεί από το GAN αυτό. Οι παράμετροι εκπαίδευσης παρουσιάζονται στον πίνακα 3.1 σύμφωνα πάντα με τους συγγραφείς του Image-Generator. Για να καταφέρει το Image-Generator να εκπαιδευτεί σωστά και να δημιουργήσει τις καλύτερες δυνατές εικόνες χρειάζεται αρχικά να εκπαιδευτεί. Όταν ολοκληρωθεί η παραπάνω διαδικασία, το Image-Generator είναι έτοιμο να δημιουργήσει νέες παραγόμενες εικόνες [51]. Ωστόσο υπάρχουν ορισμένες παράμετροι και για το στάδιο αυτό οι οποίοι παρουσιάζονται στον πίνακα 3.2.

Όνομα παραμέτρου εκπαίδευσης	Προεπιλεγμένη τιμή (Default value)	Περιγραφή
dataroot	Δεν υπάρχει προεπιλεγμένη τιμή (Απαραίτητο πεδίο – required)	Το path που βρίσκεται το σύνολο δεδομένων για την εκπαίδευση του GAN (Path of source image dataset)
netD	None	Το path για το προεκπαιδευμένο (pretrained/checkpoint) αρχείο του discriminator. Εάν δεν παρέχεται η εκπαίδευση θα ξεκινήσει από την αρχή.
netG	None	Το path για το προεκπαιδευμένο (pretrained/checkpoint) αρχείο του generator. Εάν δεν παρέχεται η εκπαίδευση θα ξεκινήσει από την αρχή.
workers	2	Ο αριθμός των workers για το dataloading
batch_size	128	Ο αριθμός του batch size για να εκπαιδευτεί το GAN
image_size	64	Ορίζει το ύψος και το πλάτος κάθε παραγόμενης εικόνας (Height-width of the generated image)
nc	3	Ο αριθμός των καναλιών της παραγόμενης εικόνας. Το 3 δηλώνει ότι η εικόνα θα είναι έγχρωμη.
nz	100	Το μέγεθος του διανύσματος latent vector z κατά την έξοδο του generator
ngf	64	Το μέγεθος των feature maps του generator
ndf	64	Το μέγεθος των feature maps του discriminator
num_epochs	5	Αριθμός εποχών εκπαίδευσης (epochs for training)
lr	0,0002	Learning Rate
beta1	0,5	Η τιμή Beta 1 για το Adam Optimizer (Beta 1 value for Adam Optimizer)
ngpu	1	Αριθμός των GPUs που θα χρησιμοποιήσει
save_every	5	Φτιάχνει ένα checkpoint μετά από κάθε save_every εποχές (epoch)
outputD	checkpoints/netD.pth	Το path όπου θα αποθηκευτεί το μοντέλο του discriminator
outputG	checkpoints/netG.pth	Το path όπου θα αποθηκευτεί το μοντέλο του generator

Πίνακας 3. 1: Παρουσιάζονται οι παράμετροι εκπαίδευσης του Image-Generator, οι default τιμές αυτών και η περιγραφή τους [52].

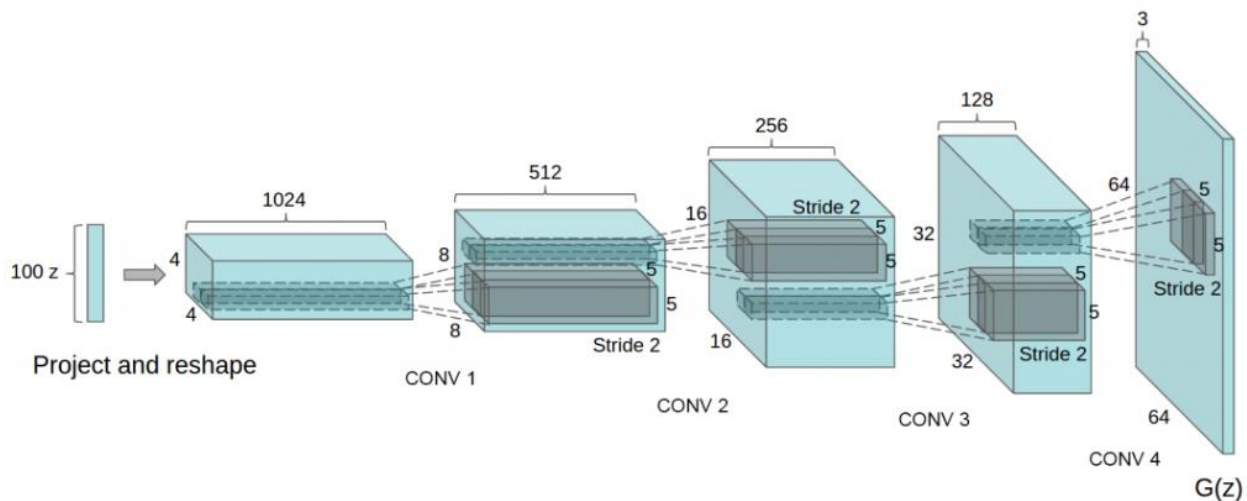
Όνομα παραμέτρου	Προεπιλεγμένη τιμή (Default value)	Περιγραφή
netG	Δεν υπάρχει προεπιλεγμένη τιμή (Απαραίτητο πεδίο – required)	Το path για το προεκπαιδευμένο (pretrained/checkpoint) αρχείο του generator το οποίο θα χρησιμοποιηθεί για την δημιουργία νέων συνθετικών εικόνων.
n	64	Ο αριθμός των εικόνων που θα δημιουργηθούν. Ωστόσο όλες οι εικόνες δεν είναι ξεχωριστές αλλά μέρος μιας εικόνας.
nc	3	Ο αριθμός των καναλιών της παραγόμενης εικόνας. Το 3 δηλώνει ότι η εικόνα θα είναι έγχρωμη.
nz	100	Το μέγεθος του διανύσματος latent vector z κατά την έξοδο του generator
ngf	64	Το μέγεθος των feature maps του generator
ngpu	1	Αριθμός των GPUs που θα χρησιμοποιήσει για τη δημιουργία της παραγόμενης εικόνας
output_path	output/fake.png	Το path όπου οι παραγόμενες εικόνες θα αποθηκευτούν

Πίνακας 3. 2: Παρουσιάζονται οι παράμετροι του Image-Generator για την δημιουργία παραγόμενων εικόνων, οι default τιμές αυτών και η περιγραφή τους [52].

3.4.3 Αρχιτεκτονική Image-Generator

Ειδικότερα το GAN αυτό υιοθετεί και τροποποιεί τρεις αποδεδειγμένες αλλαγές πάνω στις αρχιτεκτονικές των CNN. Το πρώτο είναι ότι χρησιμοποιεί εξ' ολοκλήρου συνελκτικό δίκτυο και αντικαθιστά τις ντετερμινιστικές συναρτήσεις. Η τεχνική αυτή χρησιμοποιείται στον generator ώστε να μάθει την δική της χωρική δειγματοληψία. Η δεύτερη αλλαγή που υιοθετεί είναι η τάση για την εξάλειψη των πλήρως συνδεδεμένων επιπέδων πάνω από τα συνελκτικά χαρακτηριστικά. Ένα χαρακτηριστικό παράδειγμα που ενισχύει την εξάλειψη αυτή σύμφωνα με τους συγγραφείς του “Image-Generator” είναι το “global average pooling” το οποίο έχει χρησιμοποιηθεί σε μοντέλα ταξινόμησης εικόνων. Παρατηρήθηκε ότι το “global average pooling” αύξησε την σταθερότητα του μοντέλου αλλά κατέστη αρνητικός παράγοντας για την ταχύτητα σύγκλισης. Επιτυχημένη λύση φάνηκε να είναι η διαδικασία κατά την οποία γίνεται απευθείας σύνδεση των υψηλότερων συνελκτικών χαρακτηριστικών με την είσοδο και την έξοδο και στον generator και στον discriminator. Το πρώτο στρώμα του GAN, το οποίο παίρνει ως είσοδο μια ομοιόμορφη κατανομή θορύβου Z, θα μπορούσε να ονομαστεί ως πλήρως συνδεδεμένο αφού είναι ένας πολλαπλασιασμός μήτρας. Το αποτέλεσμα αυτού ωστόσο αναδιαμορφώνεται σε ένα tensor τεσσάρων διαστάσεων και είναι η αρχή της στοίβας συνέλιξης. Για τον discriminator από την άλλη πλευρά το τελευταίο στρώμα συνέλιξης ισοπεδώνεται και στη συνέχεια τροφοδοτείται σε μία σιγμοειδής έξοδο. Αυτό που περιεγράφηκε για τον generator αναπαρίσταται στην εικόνα 3.12. Η τρίτη αλλαγή στην αρχιτεκτονική των CNN που υιοθετήθηκε από το GAN αυτό σύμφωνα με τους συγγραφείς της εργασίας είναι η κανονικοποίηση της παρτίδας (Batch Normalization). Το Batch Normalization σταθεροποιεί τη μάθηση κανονικοποιώντας την είσοδο σε κάθε μονάδα ώστε να έχει μηδενική μέση τιμή (mean) και διακύμανση μονάδων. Αυτό ουσιαστικά αντιμετωπίζει οποιαδήποτε προβλήματα πιθανόν να εμφανίζονταν λόγω κακής αρχικοποίησης. Με αυτό τον τρόπο εμποδίζεται ο generator να συμπύκνει όλα τα δείγματα σε ένα μόνο σημείο, κάτι το οποίο θα έκανε το GAN μας να αποτύχει. Επιπλέον αποφεύχθηκε να χρησιμοποιηθεί batchnorm στην έξοδο του generator και στην είσοδο του discriminator για να διατηρηθεί η σταθερότητα του μοντέλου. Με λίγα λόγια η εργασία εφάρμοσε 5 βασικούς κανόνες για να δημιουργήσει ένα σταθερό DCGAN. Αρχικά αντικατέστησε όλα τα pooling στρώματα (layers) με κλασματικά και συνελκτικά για τον discriminator. Δεύτερον χρησιμοποίησε batchnorm στον generator και στον discriminator. Τρίτον αφαίρεσε τα πλήρως συνδεδεμένα κρυφά στρώματα για βαθύτερες αρχιτεκτονικές. Τέταρτον χρησιμοποίησε το ReLU στον generator για όλα τα επίπεδα εκτός της εξόδου όπου

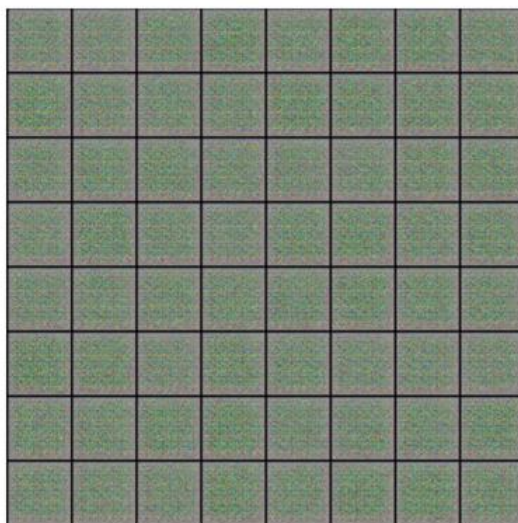
εκεί χρησιμοποιείται Tahn. Παρατηρήθηκε ότι η προσθήκη αυτή βοήθησε τον generator να εκπαιδευτεί γρηγορότερα. Τέλος χρησιμοποιήθηκε το LeakyReLU στον discriminator σε όλα τα layers αφού φάνηκε σύμφωνα με τους συγγραφείς ότι λειτουργεί καλύτερα για μοντελοποίηση υψηλότερης ανάλυσης [52][51].



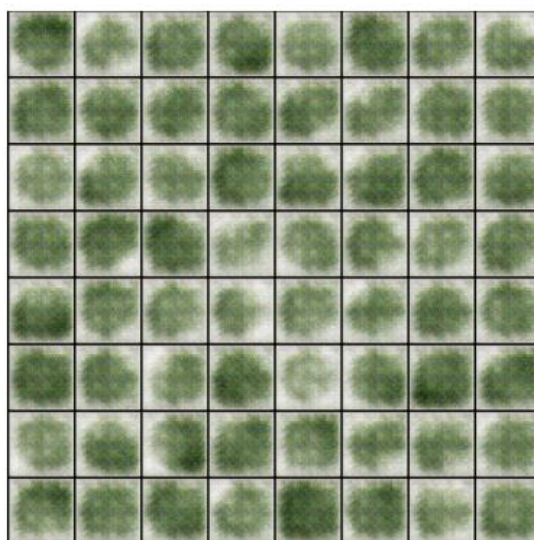
Εικόνα 3. 12: Απεικονίζεται ο generator DCGAN. Μια ομοιόμορφη κατανομή $100 z$ προβάλλεται σε μικρή χωρική έκταση συνελικτικής αναπαράστασης με πολλούς χάρτες χαρακτηριστικών. Έπειτα μια σειρά τεσσάρων κλασματικών συνελίξεων μετατρέπουν αυτήν την αναπαράσταση υψηλού επιπέδου σε εικόνα 64×64 pixel. Επίσης σημαντική παρατήρηση είναι ότι δε χρησιμοποιούνται πλήρως συνδεδεμένα (fully connected) ή pooling στρώματα. Η εικόνα πάρθηκε από [51].

3.4.4 Προσδιορισμός κατάλληλων παραμέτρων για το Image-Generator και προσαρμογή του δικτύου αυτού για την παρούσα εργασία

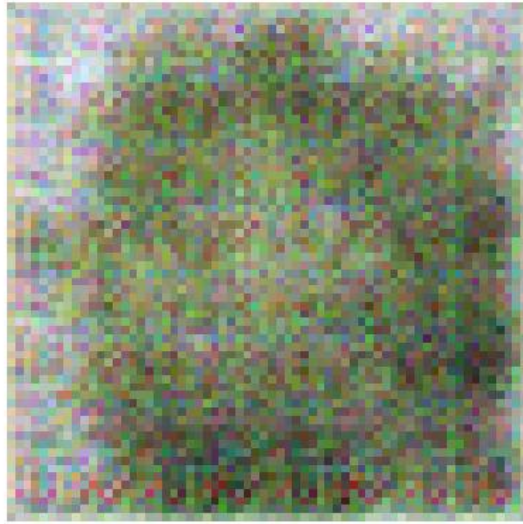
Έχοντας λοιπόν αυτό το GAN θελήσαμε να το εκπαιδύσουμε με το αντίστοιχο σύνολο δεδομένων για το πείραμα που περιγράφεται παραπάνω. Όπως ήδη έχει αναλυθεί παραπάνω η εικόνα 3.2 αποτελείται μόνο από λαχανικά και επομένως το GAN, δηλαδή το Image-Generator GAN πρέπει να εκπαιδευτεί με ένα σύνολο δεδομένων που αποτελείται μόνο από λαχανικά. Όπως αναφέρθηκε και παραπάνω θα εκπαιδευτεί με 200 εικόνες λαχανικών. Επιπλέον για το πείραμα αυτό θα χρησιμοποιήσουμε το “Google Colaboratory” ή “Colab” το οποίο είναι ένα προϊόν της Google Research. Ειδικότερα επιτρέπει σε οποιονδήποτε να γράψει και να εκτελέσει κώδικα στην Python. Εμείς χρησιμοποιούμε το εργαλείο αυτό διότι μας παρέχει την δυνατότητα να χρησιμοποιήσουμε GPU, που είναι απαραίτητη ώστε να εκτελεστεί και να τρέξει το Image-Generator. Ξεκινήσαμε λοιπόν να τρέξουμε το δίκτυο μας έχοντας σαν σύνολο εκπαίδευσης τις 200 εικόνες λαχανικών και οι τιμές των παραμέτρων του δικτύου να έχουν τις προεπιλεγμένες τιμές τους. Το αποτέλεσμα που βγάλαμε φαίνεται στην εικόνα 3.13. Όπως φαίνεται από την εικόνα απεικονίζεται μόνο θόρυβος. Θέλοντας λοιπόν να βελτιώσουμε το αποτέλεσμα θα αλλάξουμε τις τιμές κάποιων παραμέτρων. Έτσι θα εκπαιδύσουμε το δίκτυο ξανά και θα αυξήσουμε τις εποχές εκπαίδευσης μέσω της παραμέτρου “num_epochs” στις 50 ώστε να παρατηρήσουμε το αποτέλεσμα του GAN. Το αποτέλεσμα που βγάλαμε φαίνεται στην εικόνα 3.14. Όπως φαίνεται έχουν αρχίσει να σχηματίζονται εικόνες που μοιάζουν με λαχανικά. Στην συνέχεια αλλάξαμε την τιμή της παραμέτρου n και την θέσαμε ίση με 1 διότι σκοπός είναι η δημιουργία μιας παραγόμενης εικόνας που θα απεικονίζει το τι θα μπορούσε να μαγειρέψει ο χρήστης με τα ήδη υπάρχοντα υλικά του, στην προκειμένη περίπτωση με λαχανικά. Στις εικόνες 3.15, 3.16, 3.17 και 3.18 φαίνονται οι παραγόμενες εικόνες του δικτύου όταν αυτό εκπαιδευτεί σε 50, 100, 200 και 300 εποχές αντίστοιχα. Όσο αυξάνονται οι εποχές εκπαίδευσης φαίνεται να βελτιώνεται η παραγόμενη εικόνα ωστόσο συγκρίνοντας την εικόνα που δημιουργήθηκε όταν το μοντέλο εκπαιδεύτηκε στις 200 εποχές και με εκείνη που δημιουργήθηκε όταν το μοντέλο εκπαιδεύτηκε στις 300 εποχές, δεν παρατηρούμε κάποια ιδιαίτερη βελτίωση. Γι’ αυτό τον λόγο θα θεωρήσουμε ότι ιδανικά το μοντέλο μας πρέπει να εκπαιδευτεί 200 εποχές.



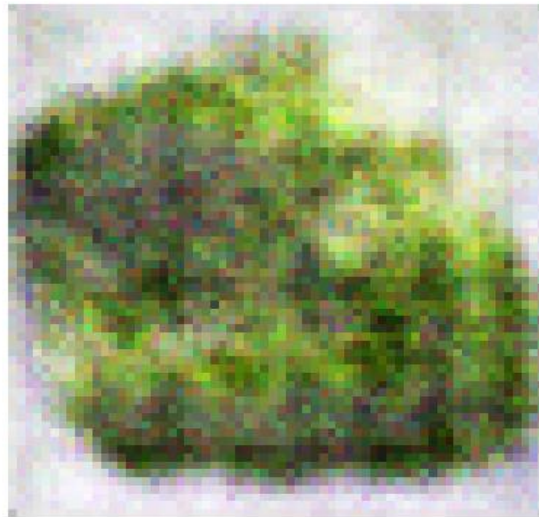
Εικόνα 3. 13: Η παραγόμενη εικόνα του Image-Generator GAN όταν χρησιμοποιούνται οι ήδη υπάρχουσες τιμές για όλες τις παραμέτρους.



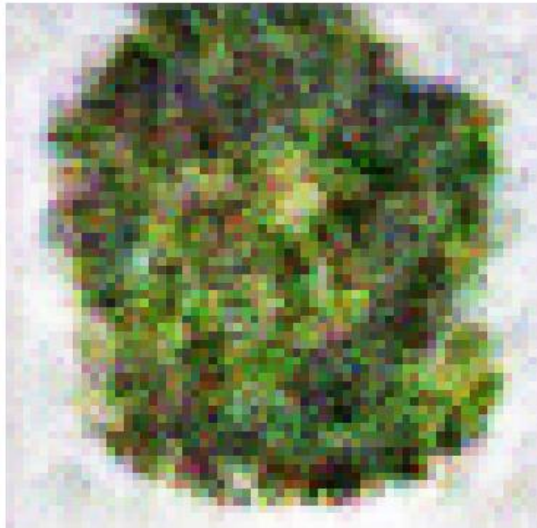
Εικόνα 3. 14: Η παραγόμενη εικόνα του Image-Generator GAN για $num_epochs = 50$. Το GAN εκπαιδεύτηκε μόνο με 200 εικόνες λαχανικών.



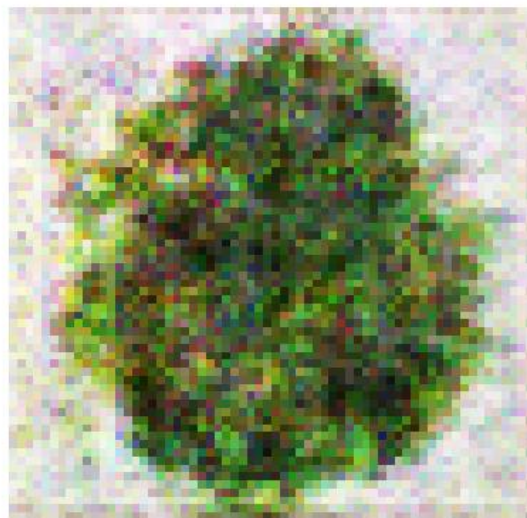
Εικόνα 3. 15: Η παραγόμενη εικόνα του Image-Generator GAN για $num_epochs = 50$ και $n = 1$ δηλαδή να δημιουργηθεί μια μόνο εικόνα. Το GAN εκπαιδεύτηκε μόνο με 200 εικόνες λαχανικών.



Εικόνα 3. 16: Η παραγόμενη εικόνα του Image-Generator GAN για $num_epochs = 100$ και $n = 1$ δηλαδή να δημιουργηθεί μια μόνο εικόνα. Το GAN εκπαιδεύτηκε μόνο με 200 εικόνες λαχανικών.

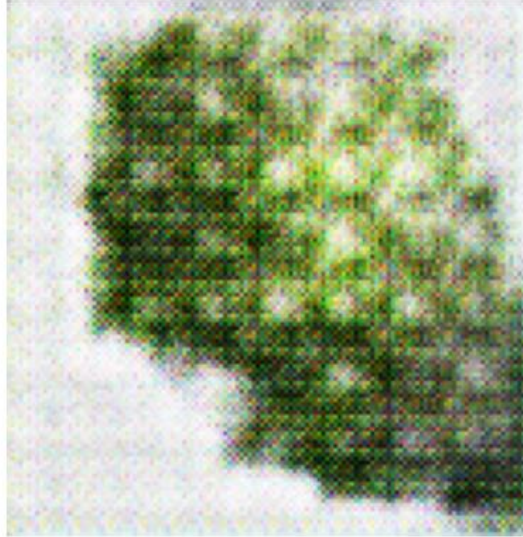


Εικόνα 3. 17: Η παραγόμενη εικόνα του Image-Generator GAN για $num_epochs = 200$ και $n = 1$ δηλαδή να δημιουργηθεί μια μόνο εικόνα. Το GAN εκπαιδεύτηκε μόνο με 200 εικόνες λαχανικών.



Εικόνα 3. 18: Η παραγόμενη εικόνα του Image-Generator GAN για $num_epochs = 300$ και $n = 1$ δηλαδή να δημιουργηθεί μια μόνο εικόνα. Το GAN εκπαιδεύτηκε μόνο με 200 εικόνες λαχανικών.

Παρατηρώντας τις εικόνες 3.17 και 3.18 φαίνεται ότι οι παραγόμενες εικόνες δεν έχουν πολύ καλή ανάλυση. Αυτό συμβαίνει γιατί εξ' ορισμού το Image-Generator δημιουργεί εικόνες με μέγεθος 64 x 64 pixels. Θα διπλασιάσουμε την τιμή της παραμέτρου "image_size" ώστε το GAN να δημιουργήσει εικόνες 128 x 128 pixels. Εκτός από τον διπλασιασμό όμως της τιμής της παραμέτρου χρειάζεται να προσθέσουμε ένα επιπλέον convolutional στρώμα (layers) και στον generator και στον discriminator. Επίσης θέσαμε την τιμή 128 για την παράμετρο "ngf" και την τιμή 32 για την παράμετρο "ndf".



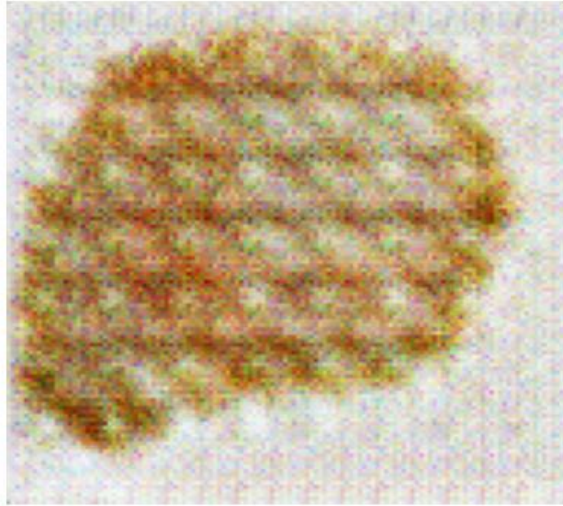
Εικόνα 3. 19: Η παραγόμενη εικόνα 128 x 128 pixels του Image-Generator GAN για $num_epochs = 200$ και $n = 1$ δηλαδή να δημιουργηθεί μια μόνο εικόνα. Το GAN εκπαιδεύτηκε μόνο με 200 εικόνες λαχανικών.

3.5 Σχολιασμός αποτελέσματος

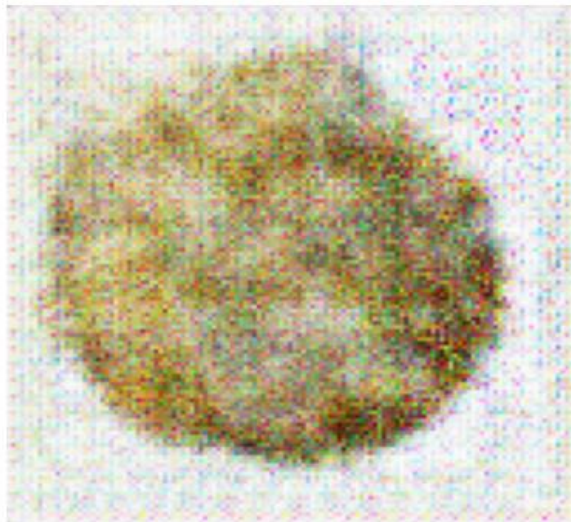
Παρατηρώντας την εικόνα 3.19 βλέπουμε ότι η ανάλυση της είναι αρκετά καλύτερη από την ανάλυση που είχαν οι προηγούμενες παραγόμενες εικόνες. Είναι φανερό ότι η εικόνα έχει αρχίσει να πλησιάζει την μορφή μιας εικόνας λαχανικών. Αυτό φαίνεται αρχικά από τα χρώματα αφού έχουν τις αποχρώσεις του πράσινου. Βέβαια ένα ανθρώπινο μάτι μπορεί εύκολα να διακρίνει ότι η εικόνα αυτή δεν είναι αληθινή αφού το δίκτυο μας δεν κατάφερε να δημιουργήσει μια εικόνα με όλες τις λεπτομέρειες που θα μπορούσε να έχει μια πραγματική εικόνα λαχανικών. Ένας από τους παράγοντες που οφείλεται αυτό είναι το μικρό μέγεθος του συνόλου των δεδομένων που χρησιμοποιήθηκαν για την εκπαίδευση του νευρωνικού δικτύου. Οι 200 εικόνες δεν είναι πολλές ώστε ένα δίκτυο να καταφέρει να εκπαιδευτεί πλήρως και να δημιουργεί ιδανικές εικόνες λαχανικών που θα μπερδευαν μέχρι και ένα ανθρώπινο μάτι. Ωστόσο το αποτέλεσμα μας είναι αρκετά ικανοποιητικό.

3.6 Παραγόμενες εικόνες από το Image-Generator

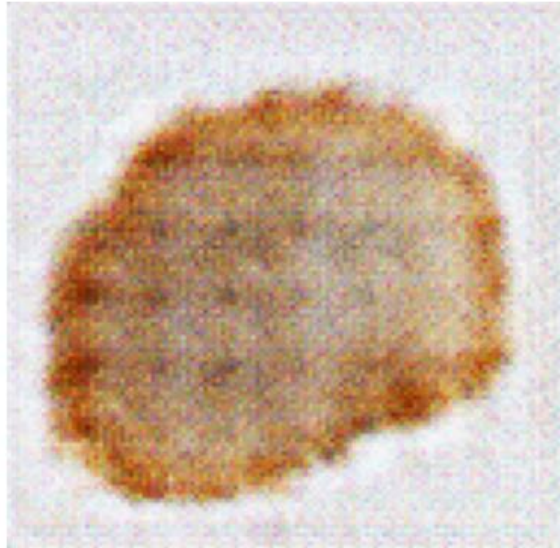
Παρακάτω παρουσιάζονται διάφορες εικόνες που έχουν παραχθεί από το Image-Generator από διαφορετικά σύνολα εκπαίδευσης. Όπως ήδη έχει αναλυθεί το σύνολο δεδομένων που εισάγεται κάθε φορά στο GAN για να εκπαιδευτεί καθορίζεται από τις ετικέτες (labels) που αναγνωρίζει το Vision API και που ανήκουν σε μία από τις έξι επιθυμητές κατηγορίες. Μπορεί στο παράδειγμα που παρουσιάζουμε η φωτογραφία να χαρακτηρίζεται μόνο από την κατηγορία “Vegetables”, υπάρχουν όμως άλλες περιπτώσεις όπου άλλες λήψεις να χαρακτηρίζονται από μία από τις άλλες πέντε κατηγορίες ή και από περισσότερες από μια κατηγορίες, δηλαδή για παράδειγμα το Vision API να επιστρέφει τις ετικέτες “Pasta” και “Potatoes”. Παρακάτω παρουσιάζονται διάφορα παραδείγματα παραγόμενων εικόνων από το GAN.



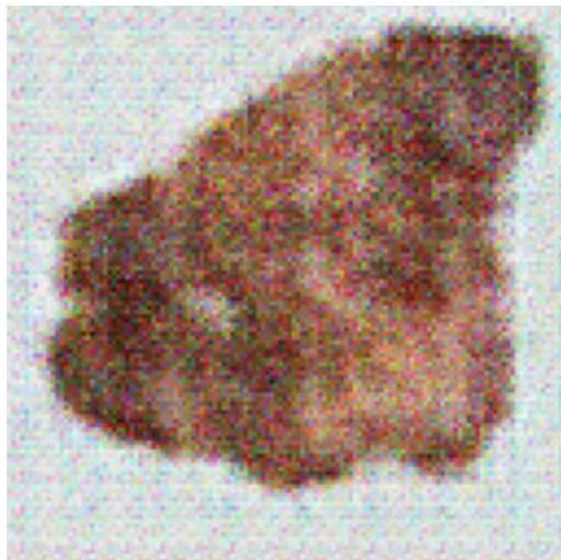
Εικόνα 3. 20: Η παραγόμενη εικόνα 128 x 128 pixels του Image-Generator GAN για num_epochs = 200 και n = 1. Το GAN εκπαιδεύτηκε μόνο με 200 εικόνες μακαρονιών.



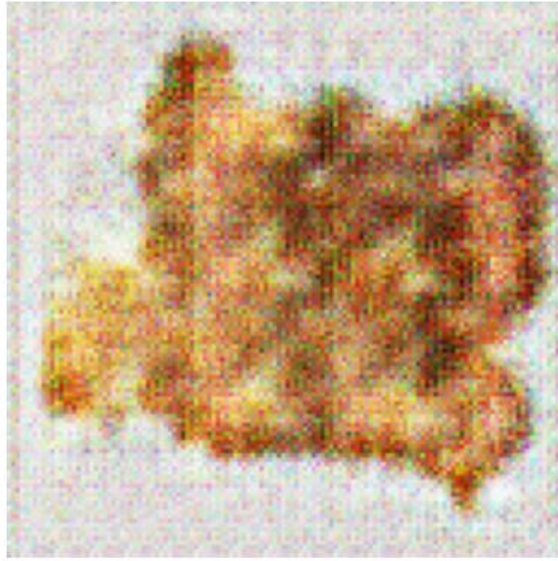
Εικόνα 3. 21: Η παραγόμενη εικόνα 128 x 128 pixels του Image-Generator GAN για num_epochs = 200 και n = 1. Το GAN εκπαιδεύτηκε μόνο με 200 εικόνες ρυζιού.



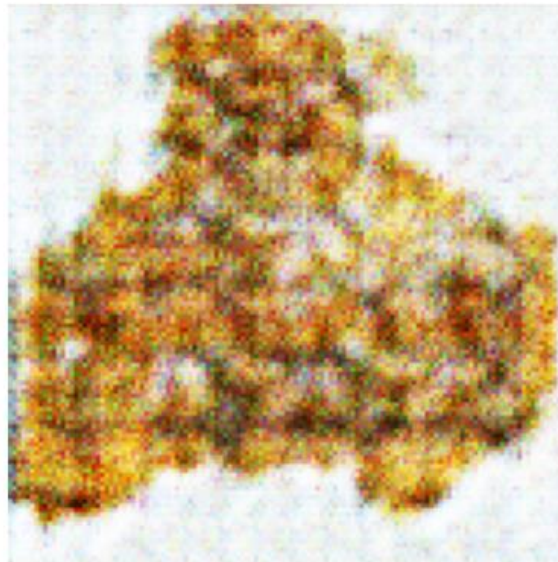
Εικόνα 3. 22: Η παραγόμενη εικόνα 128 x 128 pixels του Image-Generator GAN για num_epochs = 200 και n = 1. Το GAN εκπαιδεύτηκε μόνο με 200 εικόνες ψωμιού.



Εικόνα 3. 23: Η παραγόμενη εικόνα 128 x 128 pixels του Image-Generator GAN για num_epochs = 200 και n = 1. Το GAN εκπαιδεύτηκε μόνο με 200 εικόνες κρέατος.



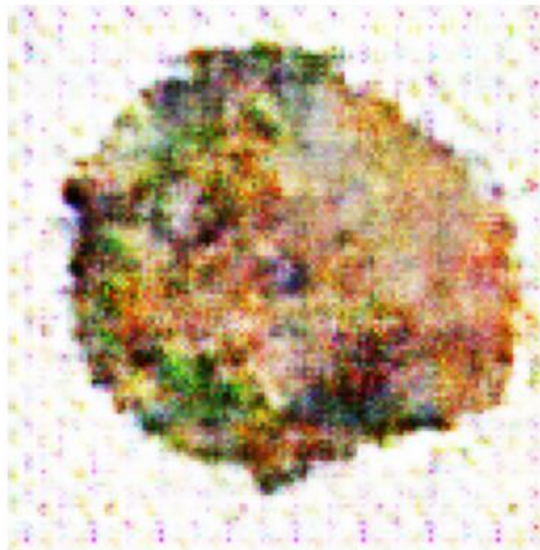
Εικόνα 3. 24: Η παραγόμενη εικόνα 128 x 128 pixels του Image-Generator GAN για num_epochs = 200 και n = 1. Το GAN εκπαιδεύτηκε μόνο με 200 εικόνες πατατών.



Εικόνα 3. 25: Η παραγόμενη εικόνα 128 x 128 pixels του Image-Generator GAN για num_epochs = 200 και n = 1. Το GAN εκπαιδεύτηκε με 200 εικόνες μακαρονιών και 200 εικόνες πατατών.



Εικόνα 3. 26: Η παραγόμενη εικόνα 128 x 128 pixels του Image-Generator GAN για $num_epochs = 200$ και $n = 1$. Το GAN εκπαιδεύτηκε με 200 εικόνες κρέατος, 200 εικόνες ρυζιού και 200 εικόνες λαχανικών.



Εικόνα 3. 27: Η παραγόμενη εικόνα 128 x 128 pixels του Image-Generator GAN για $num_epochs = 200$ και $n = 1$. Το GAN εκπαιδεύτηκε με 200 εικόνες κρέατος, 200 εικόνες ρυζιού, 200 εικόνες ψωμιού, 200 εικόνες πατατών και 200 εικόνες λαχανικών.

3.7 Σχολιασμός παραγόμενων εικόνων

Οι παραπάνω εικόνες που δημιουργήθηκαν από το GAN έχουν αρκετό ενδιαφέρον. Αρχικά παρατηρείται ότι κάποιες εικόνες διαφέρουν αρκετά μεταξύ τους ενώ κάποιες άλλες έχουν περισσότερες ομοιότητες. Συγκεκριμένα η εικόνα 3.23 διαφέρει από τις υπόλοιπες κυρίως ως προς το χρώμα. Για να

δημιουργηθεί, το νευρωνικό δίκτυο εκπαιδεύτηκε με εικόνες κρέατος οι οποίες έχουν σκούρο καφέ χρώμα, όπως και η παραγόμενη εικόνα. Αντίστοιχα και η εικόνα 3.18 που απεικονίζει λαχανικά διαφέρει αρκετά από τις υπόλοιπες. Οι εικόνες 3.20 και 3.22 που απεικονίζουν μακαρόνια και ψωμί μοιάζουν αρκετά μεταξύ τους, κάτι το οποίο είναι λογικό αφού τα χρώματα των μακαρονιών είναι παρόμοια αν όχι ίδια με τα χρώματα του ψωμιού. Επίσης και η εικόνα 3.24 που απεικονίζει πατάτες θα μπορούσαμε να πούμε ότι μοιάζει αρκετά με τις 2 προηγούμενες που αναφέραμε, παρόλα αυτά έχει πιο κιτρινωπές αποχρώσεις. Επίσης παρατηρούμε ότι και το σχήμα των εικόνων που δημιούργησε το δίκτυο παραπέμπει στο φαγητό που απεικονίζει. Για παράδειγμα 3.21 που απεικονίζει ρύζι έχει κυκλική μορφή αλλά και οι εικόνες που απεικονίζουν κρέας ή ψωμί έχουν το σωστό σχήμα. Αρκετά ικανοποιητικά είναι και τα αποτελέσματα των εικόνων που απεικονίζουν παραπάνω από μια κατηγορία φαγητού. Στην εικόνα 3.26 η οποία απεικονίζει λαχανικά, ρύζι και κρέας φαίνεται να έχει πράσινα, καφέ και μπεζ χρώματα όπως και τα υλικά από τα οποία αποτελείται. Αντίστοιχα και η εικόνα 3.27 η οποία περιέχει μεγάλη γκάμα χρωμάτων αφού αποτελείται και από όλες τις κατηγορίες εκτός εκείνης των μακαρονιών. Γενικά το νευρωνικό μας δίκτυο φαίνεται να εκπαιδεύτηκε και κατάφερε να δημιουργήσει εικόνες που πλησιάζουν στο επιθυμητό. Ωστόσο δεν έχει εκπαιδευτεί σε τέτοιο βαθμό ώστε να απεικονίζει λεπτομέρειες των φαγητών πέρα από το σχήμα και το χρώμα. Αυτό οφείλεται και στο περιορισμένο σύνολο δεδομένων πάνω στο οποίο εκπαιδεύτηκε κάθε φορά. Παρακάτω παρουσιάζεται η αξιολόγηση του δικτύου με περισσότερες λεπτομέρειες.

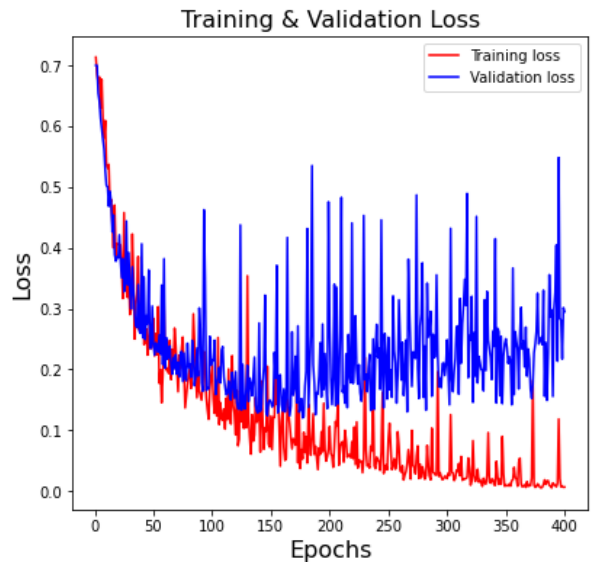
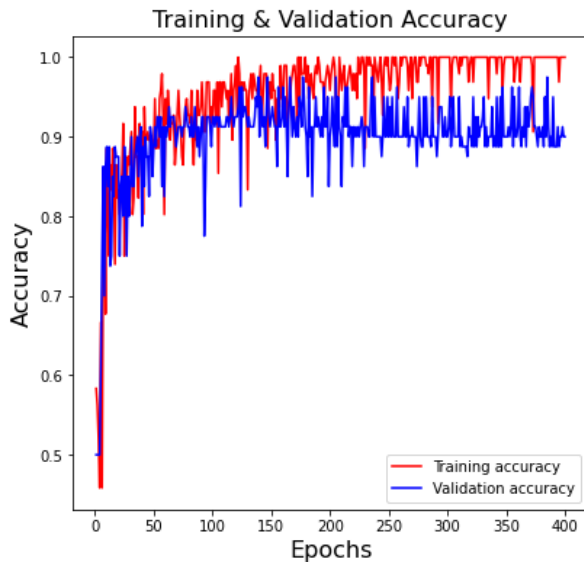
Κεφάλαιο 4: Αξιολόγηση παραγόμενων εικόνων

Για να αξιολογηθεί η ποιότητα των εικόνων που δημιουργήθηκαν από το GAN, χρησιμοποιούμε ένα Συνελικτικό Νευρωνικό Δίκτυο (Convolutional Neural Network - CNN) το οποίο καλείται να αναγνωρίσει την κατηγορία στην οποία ανήκουν οι παραγόμενες εικόνες, αφού έχει εκπαιδευτεί με πραγματικές εικόνες αυτών των κατηγοριών. Με αυτόν τον τρόπο θα γίνει αντιληπτό εάν οι παραγόμενες εικόνες του GAN “μοιάζουν” με πραγματικές εικόνες [53].

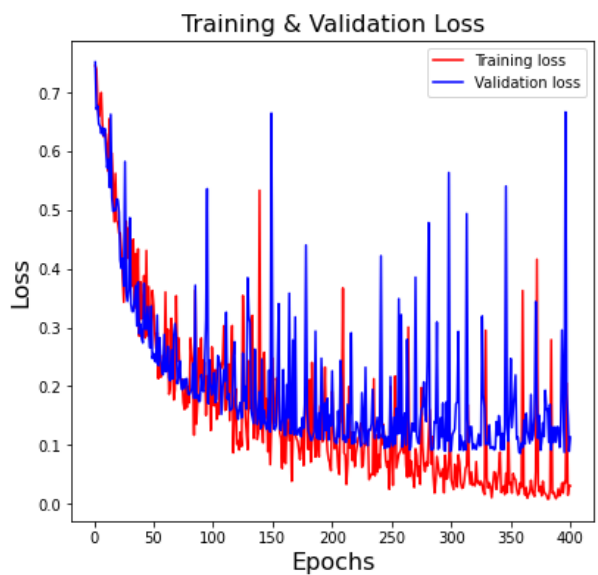
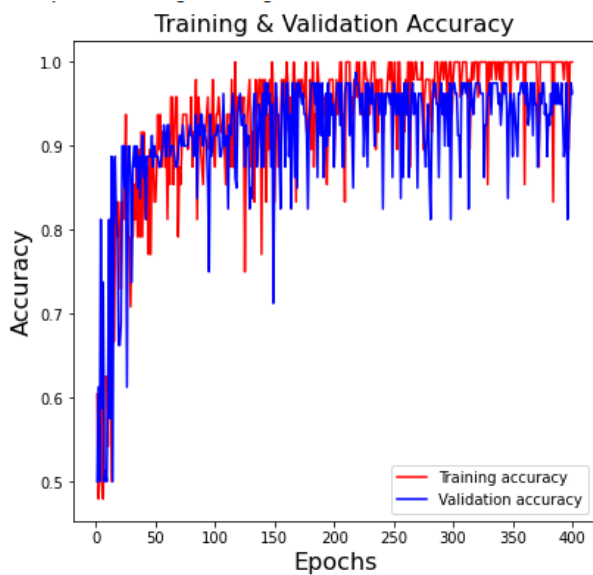
4.1 Αξιολόγηση των παραγόμενων εικόνων για τις κατηγορίες ψωμί και μακαρόνια

4.1.1 Εκπαίδευση του συνελικτικού νευρωνικού δικτύου

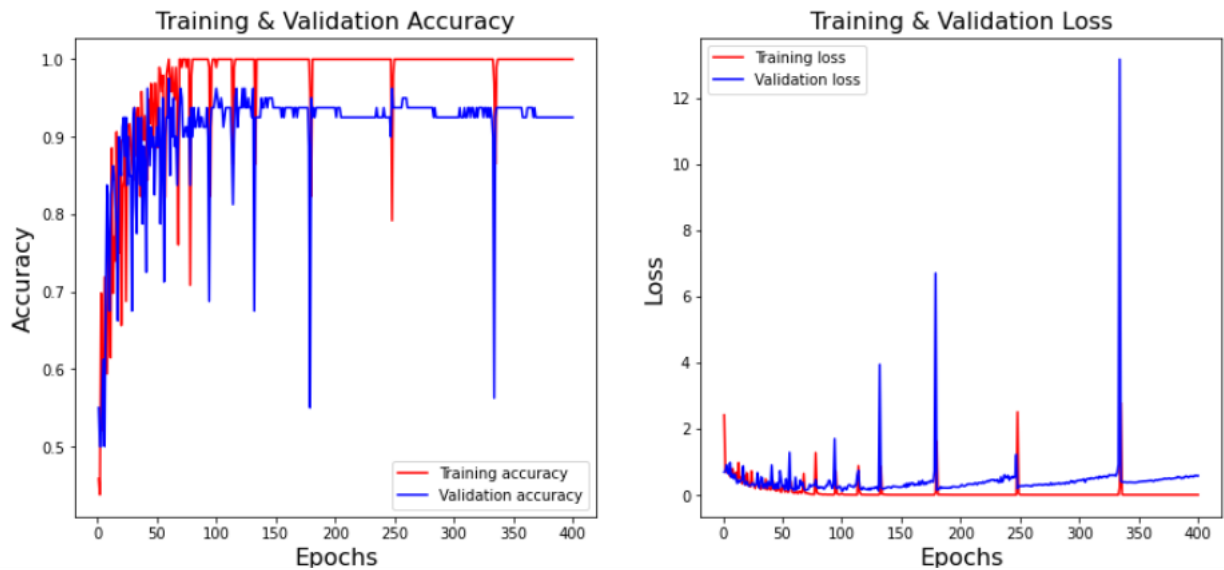
Σε αυτό το πείραμα η κλάση 1 είναι η κλάση του ψωμιού και η κλάση 2 εκείνη των μακαρονιών. Αρχικά το μοντέλο αυτό εκπαιδεύτηκε σε δύο κλάσεις εκ των οποίων η μία αποτελείται από 160 πραγματικές εικόνες ψωμιού και η δεύτερη με 160 πραγματικές εικόνες μακαρονιών. Η απόδοση του συγκεκριμένου CNN αξιολογείται με τις μετρικές Ακρίβεια (“Accuracy”) και Καμπύλη (“AUC”). Το σύνολο επικύρωσης (validation dataset) αποτελείται από 40 εικόνες ψωμιού και 40 εικόνες μακαρονιών. Το CNN εκπαιδεύτηκε αρχικά σε 400 εποχές προκειμένου να παρατηρηθεί η συνάρτηση κόστους του δικτύου (loss) αλλά και η ακρίβεια (accuracy) του. Επίσης εκπαιδεύτηκε για ρυθμό μάθησης (learning rate) ίσο με 0.001 (εικόνα 4.3) και 0.0001, όπου και παρατηρήθηκε ότι για learning rate = 0.0001 το CNN έχει καλύτερες επιδόσεις. Συγκεκριμένα, παρατηρώντας τις εικόνες 4.3 και 4.1 φαίνεται ότι η συνάρτηση κόστους έχει πολύ μεγάλες διακυμάνσεις της τάξεως του 10^1 κάτι που υποδηλώνει ότι η διαδικασία βελτιστοποίησης των παραμέτρων του δικτύου είναι ασταθής. Για τον σκοπό αυτό δοκιμάστηκε η μείωση του learning rate. Δοκιμάστηκε, επίσης, η εκπαίδευση και για διαφορετικά “batch size” (εικόνα 4.2) όπου και φάνηκε ότι για batch size ίσο με 32 το μοντέλο δίνει τα καλύτερα αποτελέσματα. Ενδεικτικά συγκρίνοντας τις εικόνες 4.2 και 4.1 φαίνεται ότι εκείνο που εκπαιδεύτηκε με batch size ίσο με 16 έχει πιο ασταθή διαγράμματα με μεγαλύτερες διακυμάνσεις. Παρακάτω παρουσιάζονται οι εικόνες με τα διαγράμματα που αποτυπώνουν την σχέση ανάμεσα στην ακρίβεια του μοντέλου (accuracy) και στις εποχές (epochs) αλλά και στις απώλειες αυτού (loss) πάλι συγκριτικά με τις εποχές (epochs) για τις διάφορες τιμές των παραμέτρων. Ειδικότερα με την κόκκινη γραμμή απεικονίζονται οι τιμές που παρατηρήθηκαν κατά την εκπαίδευση (training), ενώ με την συνεχόμενη μπλε γραμμή παρατηρείται η πορεία της “ακρίβειας του μοντέλου” και των απωλειών κατά το στάδιο επικύρωσης του μοντέλου (validation).



Εικόνα 4. 1: Διάγραμμα α) Στο αριστερά διάγραμμα απεικονίζονται οι τιμές για την ακρίβεια του μοντέλου (accuracy) από την πρώτη εποχή (1η) έως την τετρακοσιοστή (400ή) που ήταν και η τελευταία με learning rate = 0.0001 και batch = 32. β) Στο δεξιά διάγραμμα απεικονίζονται οι τιμές για τις απώλειες του μοντέλου (loss) για τις ίδιες εποχές με το διάγραμμα στα αριστερά. Η εκπαίδευση έγινε στις κλάσεις ψωμιού και μακαρονιών.



Εικόνα 4. 2: Διάγραμμα α) Στο αριστερά διάγραμμα απεικονίζονται οι τιμές για την ακρίβεια του μοντέλου (accuracy) από την πρώτη εποχή (1η) έως την τετρακοσιοστή (400ή) που ήταν και η τελευταία με learning rate = 0.0001 και batch = 16. β) Στο δεξιά διάγραμμα απεικονίζονται οι τιμές για τις απώλειες του μοντέλου (loss) για τις ίδιες εποχές με το διάγραμμα στα αριστερά. Η εκπαίδευση έγινε στις κλάσεις ψωμιού και μακαρονιών.



Εικόνα 4. 3: Διάγραμμα α) Στο αριστερά διάγραμμα απεικονίζονται οι τιμές για την ακρίβεια του μοντέλου (accuracy) από την πρώτη εποχή (1η) έως την τετρακοσιοστή (400ή) που ήταν και η τελευταία με learning rate = 0.001 και batch = 32. β) Στο δεξιά διάγραμμα απεικονίζονται οι τιμές για τις απώλειες του μοντέλου (loss) για τις ίδιες εποχές με το διάγραμμα στα αριστερά. Η εκπαίδευση έγινε στις κλάσεις ψωμιού και μακαρονιών.

Από τις παραπάνω εικόνες (εικόνες 4.1α και 4.1β) μπορούν να εξαχθούν ορισμένα συμπεράσματα. Παρατηρώντας την εικόνα 4.1β φαίνεται ότι οι απώλειες κατά την εκπαίδευση (training), αλλά και την αξιολόγηση της εκπαίδευσης (validation) αρχίζουν να ελαχιστοποιούνται μετά την 100ή εποχή, διατηρώντας φυσικά κάποιες διακυμάνσεις, οι οποίες όμως είναι μικρές της τάξεως του 10^{-1} . Ειδικότερα παρατηρώντας την συμπεριφορά της καμπύλης της εκπαίδευσης (training), όπως είναι λογικό, φαίνεται το μοντέλο συνεχώς να μειώνει τις απώλειες του. Από την εικόνα 4.1α είναι ξεκάθαρο ότι η ακρίβεια του μοντέλου τείνει να πλησιάζει το 1 που είναι και το επιθυμητό. Παρατηρώντας την μπλε γραμμή του validation φαίνεται ότι περίπου από την 150η εποχή οι απώλειες αρχίζουν και αυξάνονται αντί να μειώνονται. Παρατηρείται λοιπόν ότι το CNN δίκτυο πιθανόν να έχει υποστεί υπερπροσαρμογή (overfitting). Δηλαδή αρχίζει να χάνει την δυνατότητα σωστής ταξινόμησης των εικόνων ψωμιού και μακαρονιών. Αυτό ουσιαστικά σημαίνει ότι μέχρι την εποχή που υπερμοντελοποιείται έχει εκπαιδευτεί στο βαθμό που χρειάζεται, χωρίς να εστιάζει σε ανούσιες λεπτομέρειες οι οποίες δεν βοηθούν στην σωστή αναγνώριση των εικόνων. Επίσης από την εικόνα 4.1α φαίνεται ότι περίπου από την 120ή εποχή το μοντέλο έχει εκπαιδευτεί. Στις μετ' έπειτα εποχές δεν αυξάνει η τιμή της ακρίβειας του, δηλαδή η ικανότητα του να αναγνωρίζει κάθε εικόνα με μεγαλύτερη ακρίβεια, αλλά παραμένει πάνω κάτω ίδια και γύρω στο 0.9. Γι' αυτό λοιπόν στην συνέχεια το συγκεκριμένο νευρωνικό δίκτυο εκπαιδεύτηκε ξανά σε 100, όπου φαίνονται οι πρώτες υψηλές τιμές για το "accuracy" και οι χαμηλές για το "loss", σε 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 210 και 220 εποχές προκειμένου να παρατηρηθούν οι διαφορές στα αποτελέσματά του. Επιπλέον με αυτό τον τρόπο θα ληφθεί το συμπέρασμα για το μέχρι ποια εποχή πρέπει να εκπαιδευτεί ώστε να παράγει τα καλύτερα δυνατά αποτελέσματα. Το μοντέλο αξιολογήθηκε πάνω στο validation dataset.

Λαμβάνοντας υπόψη τα παραπάνω αποτελέσματα του CNN δικτύου για τις διάφορες εποχές εκπαίδευσης αλλά και τις κατανομές των εικόνων ανάλογα με την κλάση που ανήκουν αλλά και με εκείνη που προέβλεψε το μοντέλο λαμβάνουμε ορισμένα συμπεράσματα.

Για να υπολογίσουμε τη μήτρα σύγχυσης "confusion matrix" δηλαδή τον πίνακα που μας επιτρέπει να οπτικοποιήσουμε την απόδοση του αλγορίθμου θα θέσουμε ως "Positive" δηλαδή ως "Θετικά" τις εικόνες του ψωμιού και αντίστοιχα ως "Αρνητικά" δηλαδή ως "Negative" τις εικόνες των μακαρονιών. Συγκεκριμένα στην πρώτη στήλη του πίνακα 4.1 φαίνονται οι εποχές κατά τις οποίες εκπαιδεύτηκε κάθε φορά το μοντέλο. Στην δεύτερη στήλη του πίνακα ορίζεται ως true positive (TP) οι εικόνες ψωμιού που αναγνωρίστηκαν ως εικόνες ψωμιού. Στην τρίτη στήλη ως false negative (FN) ορίζονται οι εικόνες ψωμιού που λανθασμένα αναγνωρίστηκαν από το μοντέλο ως εικόνες μακαρονιών. Ως false positive (FP) ορίζονται οι εικόνες μακαρονιών, οι οποίες λανθασμένα αναγνωρίστηκαν ως εικόνες ψωμιού και τέλος ως true negative (TN)

ορίζονται οι εικόνες μακαρονιών που σωστά αναγνωρίστηκαν ως εικόνες μακαρονιών για την τέταρτη και πέμπτη στήλη αντίστοιχα.

Training epochs	True Positive (TP)	False Negative (FN)	False Positive (FP)	True Negative (TN)
100	39	1	6	34
110	25	15	0	40
120	31	9	0	40
130	39	1	4	36
140	35	5	0	40
150	37	3	0	40
160	36	4	1	39
170	39	1	2	38
180	39	1	2	38
190	37	3	1	39
200	38	2	1	39
210	32	8	0	40
220	39	1	2	38

Πίνακας 4. 1: Παρουσιάζονται οι τιμές για το True Positive, False Negative, False Positive και το True Negative για τις εποχές 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 210 και 220. Ως Positive ορίζονται οι εικόνες ψωμιού και ως Negative ορίζονται οι εικόνες μακαρονιών. Το μοντέλο κλήθηκε να κατηγοριοποιήσει 40 πραγματικές εικόνες ψωμιού και 40 πραγματικές εικόνες μακαρονιών (validation dataset).

Στους παρακάτω πίνακες (πίνακες 4.2 και 4.3) υπολογίζονται οι τιμές για τα True positive rate, False negative rate, False positive rate και το True negative rate αλλά και για το Accuracy, το Balanced accuracy, το Positive Predictive value, Negative predict value και το F1 score για τις ίδιες εποχές με προηγουμένως. Υπολογίζοντας αυτές τις μετρικές θα μπορέσουμε να αποφανθούμε σε ποια εποχή γίνεται η καλύτερη εκπαίδευση του CNN μοντέλου.

Training epochs	True positive rate (TPR)	False negative rate (FNR)	False positive rate (FPR)	True negative rate (TNR)
100	0.98	0.02	0.15	0.85
110	0.63	0.37	0	1
120	0.78	0.22	0	1
130	0.98	0.02	0.1	0.9
140	0.88	0.12	0	1
150	0.93	0.07	0	1
160	0.9	0.1	0.02	0.98
170	0.98	0.02	0.05	0.95
180	0.98	0.02	0.05	0.95
190	0.93	0.07	0.02	0.98
200	0.95	0.05	0.02	0.98
210	0.8	0.2	0	1
220	0.98	0.02	0.05	0.95

Πίνακας 4. 2: Παρουσιάζονται οι τιμές για το True positive rate, False negative rate, False positive rate και το True negative rate για τις εποχές 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 210 και 220. Ως Positive ορίζονται οι εικόνες ψωμιού και ως Negative ορίζονται οι εικόνες μακαρονιών. Το μοντέλο κλήθηκε να κατηγοριοποιήσει 40 πραγματικές εικόνες ψωμιού και 40 πραγματικές εικόνες μακαρονιών (validation dataset). Έγινε στρογγυλοποίηση δύο δεκαδικών ψηφίων.

Training epochs	Accuracy (ACC)	Balanced accuracy (BA)	Positive Predictive value (PPV)	Negative predict value (NPV)	F1 score
100	0.91	0.91	0.87	0.97	0.92
110	0.81	0.81	1	0.73	0.77
120	0.89	0.89	1	0.82	0.87
130	0.94	0.94	0.91	0.97	0.94
140	0.94	0.94	1	0.89	0.96
150	0.96	0.96	1	0.93	0.94
160	0.94	0.94	0.97	0.91	0.94
170	0.96	0.96	0.95	0.97	0.96
180	0.96	0.96	0.95	0.97	0.96
190	0.95	0.95	0.97	0.93	0.95
200	0.96	0.96	0.97	0.95	0.96
210	0.9	0.9	1	0.83	0.89
220	0.96	0.96	0.95	0.97	0.96

Πίνακας 4. 3: Παρουσιάζονται οι τιμές για το Accuracy, το Balanced accuracy, το Positive Predictive value, Negative predict value και το F1 score για τις εποχές 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 210 και 220. Ως Positive ορίζονται οι εικόνες ψωμιού και ως Negative ορίζονται οι εικόνες μακαρονιών. Το μοντέλο κλήθηκε να κατηγοριοποιήσει 40 πραγματικές εικόνες ψωμιού και 40 πραγματικές εικόνες μακαρονιών (validation dataset). Έγινε στρωγγυλοποίηση δύο δεκαδικών ψηφίων.

Θέλοντας να αποφανθούμε σε ποια εποχή επιτυγχάνεται η καλύτερη απόδοση και ολοκληρώνεται η εκπαίδευση, συγκρίνουμε τις τιμές του Accuracy για τις εποχές που το εκπαιδεύσαμε. Παρατηρείται ότι η μέγιστη τιμή του accuracy είναι η τιμή 0.96 και σημειώνεται στις εποχές 150, 170, 180, 200 και 220. Γι' αυτό το λόγο θα συγκριθεί η τιμή του Positive Predictive value για τις διάφορες εποχές. Φυσικά στην συγκεκριμένη περίπτωση η επιλογή της δεύτερης μετρικής είναι αυθαίρετη αφού από μόνο του το πείραμα δεν ορίζει κάποια μετρική ως σημαντικότερη μετά το Accuracy. Η επιλογή του Positive Predictive value σημαίνει πρακτικά ότι προτεραιότητα έχει το ποσοστό των εικόνων που αναγνωρίζονται ως εικόνες ψωμιού να είναι πράγματι εικόνες ψωμιού. Φαίνεται λοιπόν ότι στην 150ή εποχή το μοντέλο ολοκληρώνει την εκπαίδευσή του, αφού

είναι η μικρότερη εποχή με την μεγαλύτερη τιμή στο Accuracy αλλά και στο Positive Predictive value. Μάλιστα η τιμή του Positive Predictive value είναι 1. Αναλύοντας περισσότερο τα αποτελέσματα βλέπουμε ότι ο True positive rate (TPR) παίρνει τιμή ίση με 0.93 που σημαίνει ότι αναγνωρίζει το 93% των εικόνων του ψωμιού σωστά. Αυτόματα είναι κατανοητό ότι το False negative rate (FNR) είναι 0.07 αφού όρισε το 7% των εικόνων ψωμιού ως εικόνες μακαρονιών. Επίσης το True negative rate (TNR) είναι ίσο με 1 για τις 150 εποχές, δηλαδή αναγνωρίζει όλες τις εικόνες των μακαρονιών σωστά. Αντίστοιχα το False positive rate (FPR) είναι 0 δηλαδή καμία εικόνα μακαρονιών δεν αναγνωρίζονται ως εικόνα ψωμιού. Επιπροσθέτως το Balanced accuracy (BA) όταν το μοντέλο εκπαιδευτεί στις 150 εποχές παίρνει τιμή 0,96. Το Negative predict value (NPV) παίρνει τιμή 0,93 αφού αν το μοντέλο αναγνωρίζει μια εικόνα ως εικόνα μακαρονιών κατά 93%, δεν θα είναι εικόνα ψωμιού. Τέλος το F1 score παίρνει τιμή 0,94.

Συνοψίζοντας είναι αντιληπτό ότι το μοντέλο που εκπαιδεύτηκε σε 150 εποχές με batch size ίσο με 32 και learning rate ίσο με 0,0001, κατάφερε να αναγνωρίσει με αρκετά μεγάλη επιτυχία τις πραγματικές εικόνες ψωμιού και μακαρονιών του validation dataset. Οι προβλέψεις του μοντέλου δεν κατάφεραν να τελειοποιηθούν, δηλαδή το μοντέλο δεν κατάφερε να αναγνωρίζει και να κατηγοριοποιεί τις εικόνες ψωμιού και μακαρονιών με ακρίβεια 100% και αυτό οφείλεται στο ότι οι εικόνες αυτές μοιάζουν μεταξύ τους. Ωστόσο βρέθηκαν οι τιμές των παραμέτρων που χρειάζονται για να αναγνωρίζει και να κατηγοριοποιεί το μοντέλο με επιτυχία 96% τις πραγματικές εικόνες ψωμιού και μακαρονιών.

4.1.2 Αξιολόγηση των παραγόμενων εικόνων για τις κατηγορίες ψωμί και μακαρόνια με την βοήθεια του εκπαιδευμένου συνελκτικού νευρωνικού δικτύου.

Έχοντας λοιπόν ήδη εκπαιδεύσει το μοντέλο μας για τις κλάσεις ψωμιού και μακαρονιών σε 150 εποχές με batch size ίσο με 32 και learning rate ίσο με 0,0001, θα εισάγουμε σε αυτό τις παραγόμενες εικόνες του GAN. Η αξιολόγηση λοιπόν των παραγόμενων εικόνων θα προκύψει από την καλή ή κακή αναγνώριση αυτών, ως εικόνες ψωμιού, εάν είναι εικόνες ψωμιού, ή ως εικόνες μακαρονιών εάν φυσικά είναι εικόνες μακαρονιών. Έτσι εισάγουμε στο CNN 50 εικόνες ψωμιού παραγόμενες από το GAN και 50 εικόνες μακαρονιών που και αυτές έχουν δημιουργηθεί από το GAN. Παρακάτω παρουσιάζονται τα αποτελέσματα του μοντέλου με τον ίδιο τρόπο που έγινε και προηγουμένως.

Training epochs	True Positive (TP)	False Negative (FN)	False Positive (FP)	True Negative (TN)
150	50	0	29	21

Πίνακας 4. 4: Παρουσιάζονται οι τιμές για το True Positive, False Negative, False Positive και το True Negative για την εποχή 150. Ως Positive ορίζονται οι εικόνες ψωμιού και ως Negative ορίζονται οι εικόνες μακαρονιών. Το μοντέλο κλήθηκε να κατηγοριοποιήσει 50 παραγόμενες εικόνες ψωμιού και 50 παραγόμενες εικόνες μακαρονιών (test dataset).

Στους παρακάτω πίνακες (4.5 και 4.6) υπολογίζονται οι τιμές για τα True positive rate, False negative rate, False positive rate και το True negative rate αλλά και για το Accuracy, το Balanced accuracy, το Positive Predictive value, Negative predict value και το F1 score ώστε να αποφανθούμε κατά πόσο το μοντέλο μπόρεσε να κατηγοριοποιήσει σωστά τις παραγόμενες εικόνες.

Training epochs	True positive rate (TPR)	False negative rate (FNR)	False positive rate (FPR)	True negative rate (TNR)
150	1	0	0.58	0.42

Πίνακας 4. 5: Παρουσιάζονται οι τιμές για το True positive rate, False negative rate, False positive rate και το True negative rate για την εποχή 150. Ως Positive ορίζονται οι εικόνες ψωμιού και ως Negative ορίζονται οι εικόνες μακαρονιών. Το μοντέλο κλήθηκε να κατηγοριοποιήσει 50 παραγόμενες εικόνες ψωμιού και 50 παραγόμενες εικόνες μακαρονιών (test dataset). Έγινε στρογγυλοποίηση δύο δεκαδικών ψηφίων.

Training epochs	Accuracy (ACC)	Balanced accuracy (BA)	Positive Predictive value (PPV)	Negative predict value (NPV)	F1 score
150	0.71	0.71	0.63	1	0.78

Πίνακας 4. 6: Παρουσιάζονται οι τιμές για το Accuracy, το Balanced accuracy, το Positive Predictive value, Negative predict value και το F1 score για την εποχή 150. Ως Positive ορίζονται οι εικόνες ψωμιού και ως Negative ορίζονται οι εικόνες μακαρονιών. Το μοντέλο κλήθηκε να κατηγοριοποιήσει 50 παραγόμενες εικόνες ψωμιού και 50 παραγόμενες εικόνες μακαρονιών (test dataset). Έγινε στρογγυλοποίηση δύο δεκαδικών ψηφίων.

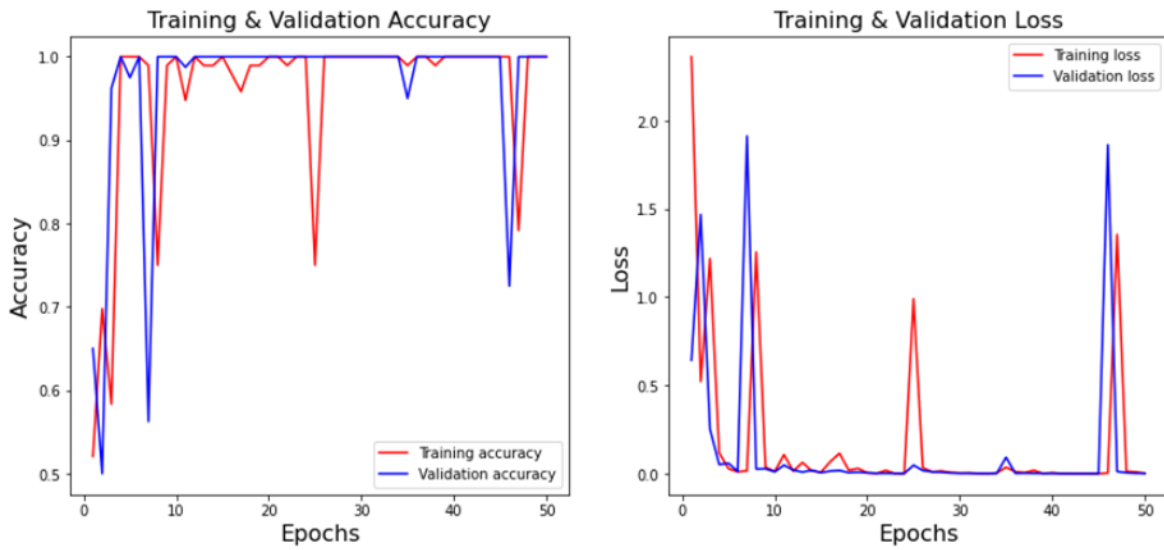
Καταλήγουμε λοιπόν ότι το μοντέλο, στις 120 εποχές και με τις κατάλληλες τιμές στις παραμέτρους, κατάφερε να αναγνωρίσει με σχετικά μεγάλη επιτυχία τις παραγόμενες εικόνες ψωμιού και μακαρονιών του GAN αυτής της εργασίας. Συγκεκριμένα η ακρίβεια (Accuracy) στην ταξινόμηση των εικόνων έφτασε το 71%. Η τιμή του Positive Predictive value (PPV) έφτασε το 0.63 δηλαδή κατά 63% εάν το μοντέλο προβλέψει ότι μια εικόνα είναι εικόνα ψωμιού, η συγκεκριμένη εικόνα είναι πράγματι εικόνα ψωμιού. Το Negative predict value (NPV) πήρε τιμή 1 αφού αν το μοντέλο αναγνωρίζει μια εικόνα ως εικόνα μακαρονιών σίγουρα η εικόνα αυτή δεν θα είναι εικόνα μακαρονιών. Τέλος και το F1 score παίρνει τιμή 0.78.

Συγκρίνοντας τις τιμές αυτές με τις αντίστοιχες τιμές για το πείραμα που εκτελέστηκε στις ίδιες εποχές και παραμέτρους, αλλά αξιολογήθηκε με το validation dataset, παρατηρούμε ορισμένες διαφορές. Συγκεκριμένα το μοντέλο που αξιολογήθηκε με το validation dataset κατάφερε να κατηγοριοποιήσει όλες τις πραγματικές εικόνες με επιτυχία 96% δηλαδή έχοντας Accuracy ίσο με 96%. Το ότι το μοντέλο δεν κατάφερε να φτάσει την απόλυτη ακρίβεια και να αναγνωρίζει τις πραγματικές εικόνες με 100% επιτυχία, φανερώνει ότι οι εικόνες ψωμιού και μακαρονιών έχουν κοινά στοιχεία που δυσκολεύουν το CNN μοντέλο να τις αναγνωρίζει με απόλυτη σιγουριά. Από την άλλη πλευρά όταν το μοντέλο κλήθηκε να ταξινομήσει τις παραγόμενες εικόνες του GAN τα κατάφερε με 71% επιτυχία. Αυτό οφείλεται στο ότι οι παραγόμενες εικόνες, όπως είναι λογικό, είναι χαμηλότερης ποιότητας από της πραγματικές, κάτι που δυσκολεύει το CNN μοντέλο να επιτύχει το ίδιο μεγάλο σκορ στην ταξινόμηση των παραγόμενων εικόνων. Ωστόσο καταφέρνει να κατηγοριοποιήσει σε ένα αρκετά μεγάλο ποσοστό (71%) εικόνες που μοιάζουν αρκετά και που ακόμα και ένα ανθρώπινο μάτι θα δυσκολευόταν να τις ξεχωρίζει με απόλυτα ακρίβεια. Βλέποντας κανείς πόσο όμοιες είναι οι παραγόμενες εικόνες ψωμιού και μακαρονιών αντιλαμβάνεται ότι το 71% είναι ένα αρκετά καλό ποσοστό ταξινόμησης. Επίσης αυτό σημαίνει ότι το GAN δημιούργησε εικόνες αρκετά καλές που ένα CNN μοντέλο εκπαιδευμένο σε πραγματικές εικόνες ψωμιού και μακαρονιών μπόρεσε να τις αναγνωρίσει. Φυσικά επειδή το επιθυμητό θα ήταν να έχει την ικανότητα να φτάσει το ποσοστό που πέτυχε το μοντέλο αναγνωρίζοντας πραγματικές εικόνες, δηλαδή να αναγνωρίζει τις παραγόμενες εικόνες με επιτυχία 96%, το GAN φαίνεται να έχει περιθώρια βελτίωσης.

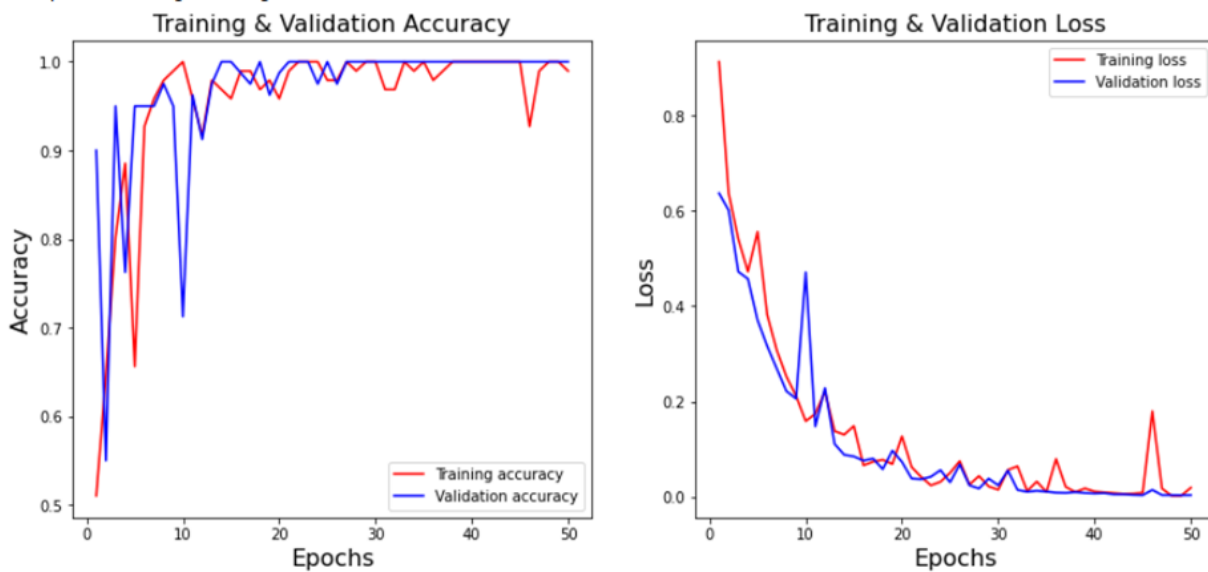
4.2 Αξιολόγηση των παραγόμενων εικόνων για τις κατηγορίες ψωμί και λαχανικών.

4.2.1 Εκπαίδευση του συνελκτικού νευρωνικού δικτύου

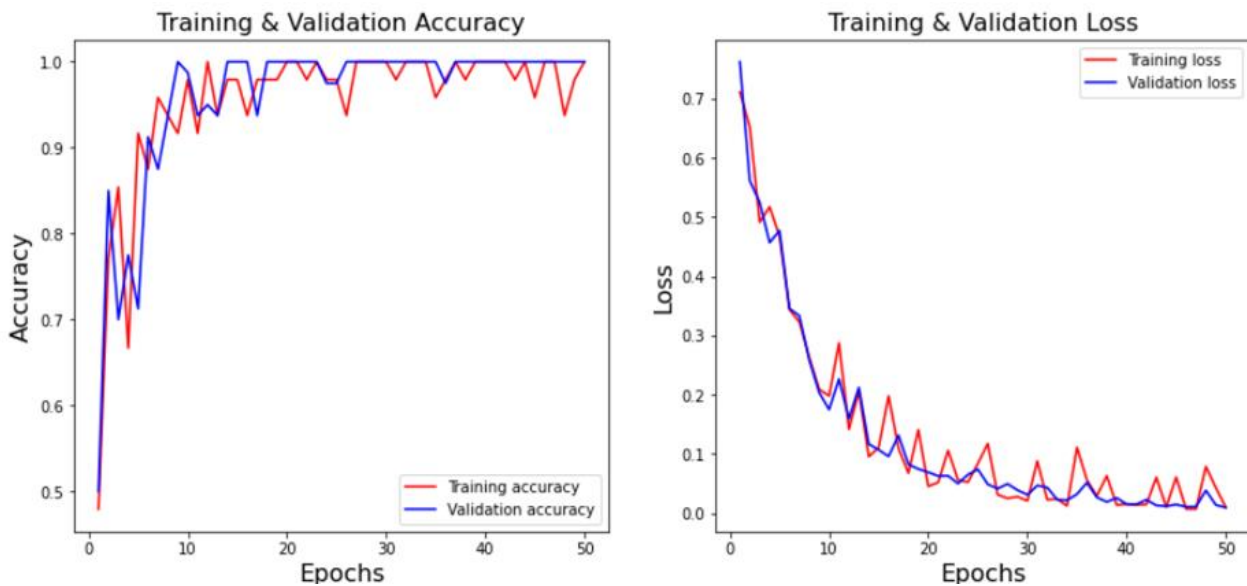
Ακριβώς η ίδια διαδικασία πραγματοποιήθηκε και για έναν άλλον συνδυασμό κλάσεων. Είναι εμφανές ότι οι παραγόμενες εικόνες του ψωμιού και των μακαρονιών είναι αρκετά παρόμοιες και για αυτό τον λόγο πραγματοποιήθηκε η αξιολόγηση και στις κλάσεις του ψωμιού με εκείνες των λαχανικών. Η επιλογή αυτή έγινε διότι οι παραγόμενες εικόνες ψωμιού διαφέρουν περισσότερο από εκείνες των μακαρονιών συγκριτικά πάντα με τις παραγόμενες εικόνες ψωμιού που θα κληθεί και το CNN δίκτυο να αναγνωρίσει και να κατηγοριοποιήσει. Στόχος είναι να παρατηρηθεί η ακρίβεια και γενικότερα τα αποτελέσματα στην κατηγοριοποίηση παραγόμενων εικόνων που δεν μοιάζουν οπτικά αρκετά μεταξύ τους, όπως οι κλάσεις ψωμιού και λαχανικών, και εικόνων που μοιάζουν περισσότερο, όπως η κλάσεις ψωμιού και μακαρονιών που παρουσιάστηκε παραπάνω. Έτσι το μοντέλο εκπαιδεύτηκε σε δύο κλάσεις εκ των οποίων η μία αποτελείται από 160 πραγματικές εικόνες ψωμιού και η δεύτερη με 160 πραγματικές εικόνες λαχανικών. Το συγκεκριμένο CNN και εδώ αξιολογείται με τις μετρικές “Accuracy” και “AUC”. Στην συνέχεια αξιολογήθηκε η εκπαίδευση που έγινε μέσω του validation dataset, το οποίο αποτελείται από 40 πραγματικές εικόνες ψωμιού και 40 πραγματικές εικόνες λαχανικών. Το συγκεκριμένο νευρωνικό δίκτυο εκπαιδεύτηκε αρχικά σε 50 εποχές προκειμένου να παρατηρηθούν οι απώλειες του δικτύου (loss) αλλά και η ακρίβεια να αναγνωρίζει τις εικόνες (accuracy) κατά την διάρκεια των εποχών. Σε σχέση με το προηγούμενο μοντέλο που εκπαιδεύτηκε να ξεχωρίζει εικόνες μακαρονιών και ψωμιού, αυτό το μοντέλο έχει “ευκολότερη” εργασία να εκτελέσει αφού ακόμα και ένα ανθρώπινο μάτι μπορεί με πολύ μεγαλύτερη ευκολία να διακρίνει τις εικόνες λαχανικών από εκείνες του ψωμιού, σε σχέση με εκείνες των μακαρονιών συγκριτικά με τις εικόνες ψωμιού. Βασικός λόγος που η κατηγοριοποίηση αυτή είναι πιο απλή είναι λόγω του χρώματος ανάμεσα στο ψωμί και τα λαχανικά. Γι’ αυτό παρατηρείται ότι το συγκεκριμένο νευρωνικό δίκτυο αρκεί να εκπαιδευτεί ως τις 50 εποχές ώστε να έχει ολοκληρωθεί η εκπαίδευση του ήδη από προηγούμενη εποχή. Επίσης εκπαιδεύτηκε για learning rate ίσο με 0.001 (εικόνα 4.4) και 0.0001 (εικόνα 4.5), όπου και παρατηρήθηκε ότι για learning rate = 0.0001 το CNN εκπαιδεύεται καλύτερα. Συγκεκριμένα παρατηρώντας τις εικόνες 4.4 και 4.5 φαίνεται στην εικόνα 4.4 ότι μετά την εποχή 40 το διάγραμμα αρχίζει να χαλάει αφού εμφανίζεται μια διακύμανση της τάξεως της μίας μονάδας. Δοκιμάστηκε η εκπαίδευση και για διαφορετικά “batch size” (εικόνα 4.5) και για batch size ίσο με 32 φαίνεται το μοντέλο αυτό να δίνει τα καλύτερα αποτελέσματα. Ωστόσο επειδή η ταξινόμηση που έχει να εκτελέσει το μοντέλο είναι σχετικά εύκολη οι διαφορές μεταξύ των διαφορετικών batch sizes δεν είναι μεγάλες. Ενδεικτικά συγκρίνοντας τις εικόνες 4.5 και 4.6 φαίνεται ότι εκείνο που εκπαιδεύτηκε με batch size ίσο με 16 έχει πιο ασταθή διαγράμματα με μεγαλύτερες διακυμάνσεις. Συγκεκριμένα κατά την αξιολόγηση της εκπαίδευσης παρατηρείται ότι όταν το batch size είναι ίσο με 16, η ακρίβεια του μοντέλου δεν μένει σταθερή στην τιμή 1 όπως όταν το batch size είναι ίσο με 32 αλλά και οι απώλειες του παίρνουν μικρότερες τιμές. Παρακάτω παρουσιάζονται οι εικόνες με τα διάγραμμα που αποτυπώνουν την σχέση ανάμεσα στην ακρίβεια του μοντέλου (accuracy) και στις εποχές (epochs) αλλά και στις απώλειες αυτού (loss) πάλι συγκριτικά με τις εποχές (epochs). Ειδικότερα με την κόκκινη γραμμή απεικονίζονται οι τιμές που παρατηρήθηκαν κατά την εκπαίδευση (training), ενώ με την συνεχόμενη μπλε γραμμή παρατηρείται η πορεία της “ακρίβειας του μοντέλου” και των απωλειών κατά την αξιολόγηση της εκπαίδευσης του μοντέλου (validation).



Εικόνα 4. 4: Διάγραμμα α) Στο αριστερά διάγραμμα απεικονίζονται οι τιμές για την ακρίβεια του μοντέλου (accuracy) από την πρώτη εποχή (1η) έως την πενήκοστη (50ή) που ήταν και η τελευταία με learning rate = 0.001 και batch = 32. β) Στο δεξιά διάγραμμα απεικονίζονται οι τιμές για τις απώλειες του μοντέλου (loss) για τις ίδιες εποχές με το διάγραμμα στα αριστερά. Η εκπαίδευση έγινε στις κλάσεις ψωμιού και λαχανικών.



Εικόνα 4. 5: Διάγραμμα α) Στο αριστερά διάγραμμα απεικονίζονται οι τιμές για την ακρίβεια του μοντέλου (accuracy) από την πρώτη εποχή (1η) έως την πενήκοστη (50ή) που ήταν και η τελευταία με learning rate = 0.0001 και batch = 32. β) Στο δεξιά διάγραμμα απεικονίζονται οι τιμές για τις απώλειες του μοντέλου (loss) για τις ίδιες εποχές με το διάγραμμα στα αριστερά. Η εκπαίδευση έγινε στις κλάσεις ψωμιού και λαχανικών.



Εικόνα 4. 6: Διάγραμμα α) Στο αριστερά διάγραμμα απεικονίζονται οι τιμές για την ακρίβεια του μοντέλου (accuracy) από την πρώτη εποχή (1η) έως την πενήκοστή (50ή) που ήταν και η τελευταία με learning rate = 0.0001 και batch = 16. β) Στο δεξιά διάγραμμα απεικονίζονται οι τιμές για τις απώλειες του μοντέλου (loss) για τις ίδιες εποχές με το διάγραμμα στα αριστερά. Η εκπαίδευση έγινε στις κλάσεις ψωμιού και λαχανικών.

Από τις παραπάνω εικόνες (εικόνες 4.5α και 4.5β) μπορούν να ληφθούν ορισμένα συμπεράσματα. Παρατηρώντας το διάγραμμα της εικόνας 4.5β φαίνεται ότι οι απώλειες κατά την εκπαίδευση (training), αλλά και την αξιολόγηση της εκπαίδευσης (validation) αρχίζουν να ελαχιστοποιούνται από την 20ή εποχή, διατηρώντας φυσικά κάποιες πολύ μικρές διακυμάνσεις. Ειδικότερα παρατηρώντας την συμπεριφορά της γραμμής της εκπαίδευσης (training), όπως είναι λογικό, φαίνεται το μοντέλο συνεχώς να μειώνει τις απώλειες του και μετά την 30ή εποχή είναι σχεδόν μηδέν με μικρές διακυμάνσεις. Από το διάγραμμα της εικόνας 4.5α είναι ξεκάθαρο ότι συνεχώς η ακρίβεια του μοντέλου πλησιάζει το 1 με μικρές διακυμάνσεις. Το ίδιο συμβαίνει και για την γραμμή της εκπαίδευσης από την εποχή 30 και μετά. Παρατηρώντας την μπλε γραμμή του validation φαίνεται ότι περίπου από την 35η εποχή οι απώλειες μειώνονται και είναι σχεδόν μηδέν. Επίσης η γραμμή του validation φαίνεται να έχει πιάσει την μέγιστη τιμή για την ακρίβεια του μοντέλου μετά την 25η εποχή. Γι' αυτό λοιπόν στην συνέχεια το συγκεκριμένο νευρωνικό δίκτυο εκπαιδεύτηκε ξανά σε 10, 15, 20, 25, 30, 35, 40 και 45 εποχές προκειμένου να παρατηρηθούν οι διαφορές στα αποτελέσματά του. Επιπλέον με αυτό τον τρόπο θα ληφθεί το συμπέρασμα για το μέχρι ποια εποχή πρέπει να εκπαιδευτεί ώστε να παράγει τα καλύτερα δυνατά αποτελέσματα. Το μοντέλο αξιολογήθηκε πάνω στο validation dataset.

Λαμβάνοντας υπόψη τα παραπάνω αποτελέσματα του CNN δικτύου για τις διάφορες εποχές εκπαίδευσης αλλά και τις κατανομές των εικόνων ανάλογα με την κλάση που ανήκουν αλλά και με εκείνη που προέβλεψε το μοντέλο λαμβάνουμε ορισμένα συμπεράσματα.

Για να υπολογίσουμε τον "confusion matrix" δηλαδή τον πίνακα που μας επιτρέπει να οπτικοποιήσουμε την απόδοση του αλγορίθμου θα θέσουμε ως "Positive" δηλαδή ως "Θετικά" τις εικόνες του ψωμιού και αντίστοιχα ως "Αρνητικά" δηλαδή ως "Negative" τις εικόνες των λαχανικών. Συγκεκριμένα στην πρώτη στήλη του πίνακα 4.7 φαίνονται οι εποχές κατά τις οποίες εκπαιδεύτηκε κάθε φορά το μοντέλο. Στην δεύτερη στήλη του πίνακα ορίζεται ως true positive (TP) οι εικόνες ψωμιού που αναγνωρίστηκαν ως εικόνες ψωμιού. Στην τρίτη στήλη ως false negative (FN) ορίζονται οι εικόνες ψωμιού που λανθασμένα αναγνωρίστηκαν από το μοντέλο ως εικόνες λαχανικών. Ως false positive (FP) ορίζονται οι εικόνες λαχανικών, οι οποίες λανθασμένα αναγνωρίστηκαν ως εικόνες ψωμιού και τέλος ως true negative (TN) ορίζονται οι εικόνες λαχανικών που σωστά αναγνωρίστηκαν ως εικόνες λαχανικών για την τέταρτη και πέμπτη στήλη αντίστοιχα.

Training epochs	True (TP)	Positive	False (FN)	Negative	False Positive (FP)	True (TN)	Negative
10	38		2		0	40	
15	39		1		1	39	
20	40		0		0	40	
25	40		0		0	40	
30	40		0		0	40	
35	40		0		0	40	
40	40		0		0	40	
45	40		0		0	40	
50	40		0		1	39	

Πίνακας 4. 7: Παρουσιάζονται οι τιμές για το True Positive, False Negative, False Positive και το True Negative για τις εποχές 10, 15, 20, 25, 30, 35, 40, 45 και 50. Ως Positive ορίζονται οι εικόνες ψωμιού και ως Negative ορίζονται οι εικόνες λαχανικών. Το μοντέλο κλήθηκε να κατηγοριοποιήσει 40 πραγματικές εικόνες ψωμιού και 40 πραγματικές εικόνες λαχανικών (validation dataset).

Στους παρακάτω πίνακες (πίνακες 4.8 και 4.9) υπολογίζονται οι τιμές για τα True positive rate, False negative rate, False positive rate και το True negative rate αλλά και για το Accuracy, το Balanced accuracy, το Positive Predictive value, Negative predict value και το F1 score για τις ίδιες εποχές με προηγουμένως. Υπολογίζοντας αυτές τις μετρικές θα μπορέσουμε να αποφανθούμε σε ποια εποχή γίνεται η καλύτερη εκπαίδευση του CNN μοντέλου.

Training epochs	True positive rate (TPR)	False negative rate (FNR)	False positive rate (FPR)	True negative rate (TNR)
10	0.95	0.05	0	1
15	0.98	0.02	0.02	0.98
20	1	0	0	1
25	1	0	0	1
30	1	0	0	1
35	1	0	0	1
40	1	0	0	1
45	1	0	0	1
50	1	0	0.02	0.98

Πίνακας 4. 8: Παρουσιάζονται οι τιμές για το True positive rate, False negative rate, False positive rate και το True negative rate για τις εποχές 10, 15, 20, 25, 30, 35, 40, 45 και 50. Ως Positive ορίζονται οι εικόνες ψωμιού και ως Negative ορίζονται οι εικόνες λαχανικών. Το μοντέλο κλήθηκε να κατηγοριοποιήσει 40 πραγματικές εικόνες ψωμιού και 40 πραγματικές εικόνες λαχανικών (validation dataset). Έγινε στρογγυλοποίηση δύο δεκαδικών ψηφίων.

Training epochs	Accuracy (ACC)	Balanced accuracy (BA)	Positive Predictive value (PPV)	Negative predict value (NPV)	F1 score
10	0.98	0.98	1	0.95	0.97
15	0.98	0.98	0.98	0.98	0.98
20	1	1	1	1	1
25	1	1	1	1	1
30	1	1	1	1	1
35	1	1	1	1	1
40	1	1	1	1	1
45	1	1	1	1	1
50	0.99	0.99	0.98	1	0.99

Πίνακας 4. 9: Παρουσιάζονται οι τιμές για το Accuracy, το Balanced accuracy, το Positive Predictive value, Negative predict value και το F1 score για τις εποχές 10, 15, 20, 25, 30, 35, 40, 45 και 50. Ως Positive ορίζονται οι εικόνες ψωμιού και ως Negative ορίζονται οι εικόνες λαχανικών. Το μοντέλο κλήθηκε να κατηγοριοποιήσει 40 πραγματικές εικόνες ψωμιού και 40 πραγματικές εικόνες λαχανικών (validation dataset). Έγινε στρωγγυλοποίηση δύο δεκαδικών ψηφίων.

Καταλήγουμε λοιπόν ότι η καλύτερη εκπαίδευση του CNN μοντέλου γίνεται όταν εκπαιδεύεται σε 20 εποχές αφού είναι η πρώτη εποχή που η τιμή του Accuracy (ACC) και του Balanced accuracy (BA) παίρνουν την τιμή 1 δηλαδή αναγνωρίζουν σωστά όλες τις εικόνες. Συγκεκριμένα το True positive rate (TPR) παίρνει τιμή ίση με το 1 που σημαίνει ότι αναγνωρίζει όλες τις εικόνες του ψωμιού σωστά. Αυτόματα είναι κατανοητό ότι το False negative rate (FNR) είναι 0 αφού δεν όρισε καμία εικόνα ψωμιού ως εικόνα λαχανικών. Επίσης το True negative rate (TNR) παίρνει και αυτό τιμή ίση με το 1 που σημαίνει ότι αναγνωρίζει όλες τις εικόνες λαχανικών σωστά. Αντίστοιχα το False positive rate (FPR) είναι 0. Το Positive Predictive value (PPV) είναι ίσο με 1 δηλαδή κατά 100% εάν το μοντέλο προβλέψει ότι μια εικόνα είναι εικόνα ψωμιού, η συγκεκριμένη εικόνα είναι πράγματι εικόνα ψωμιού. Το Negative predict value (NPV) παίρνει τιμή 1 αφού αν το μοντέλο αναγνωρίζει μια εικόνα ως εικόνα λαχανικών σίγουρα η εικόνα αυτή δεν θα είναι εικόνα ψωμιού. Τέλος και το F1 score παίρνει τιμή 1.

Συνοψίζοντας είναι αντιληπτό ότι το μοντέλο που εκπαιδεύτηκε σε 20 εποχές με batch size ίσο με 32 και learning rate ίσο με 0,0001, κατάφερε να αναγνωρίσει με 100% επιτυχία τις πραγματικές εικόνες ψωμιού και λαχανικών του validation dataset. Επίσης βρέθηκαν οι τιμές των παραμέτρων που χρειάζονται για να αναγνωρίζει και να κατηγοριοποιεί το μοντέλο με επιτυχία τις πραγματικές εικόνες ψωμιού και λαχανικών.

4.2.2 Αξιολόγηση των παραγόμενων εικόνων για τις κατηγορίες ψωμί και λαχανικών με την βοήθεια του εκπαιδευμένου συνελκτικού δικτύου.

Έχοντας λοιπόν ήδη εκπαιδεύσει το μοντέλο μας για τις κλάσεις ψωμιού και λαχανικών σε 20 εποχές με batch size ίσο με 32 και learning rate ίσο με 0,0001, θα εισάγουμε σε αυτό τις παραγόμενες εικόνες του GAN. Η αξιολόγηση λοιπόν των παραγόμενων εικόνων θα προκύψει από την καλή ή κακή αναγνώριση αυτών, ως εικόνες ψωμιού, εάν είναι εικόνες ψωμιού, ή ως εικόνες λαχανικών εάν φυσικά είναι εικόνες λαχανικών. Έτσι εισάγουμε στο CNN 50 εικόνες ψωμιού παραγόμενες από το GAN και 50 εικόνες λαχανικών που και αυτές έχουν δημιουργηθεί από το GAN. Παρακάτω παρουσιάζονται τα αποτελέσματα του μοντέλου με τον ίδιο τρόπο που έγινε και προηγουμένως.

Training epochs	True Positive (TP)	False Negative (FN)	False Positive (FP)	True Negative (TN)
20	50	0	13	37

Πίνακας 4. 10: Παρουσιάζονται οι τιμές για το True Positive, False Negative, False Positive και το True Negative για την εποχή 20. Ως Positive ορίζονται οι εικόνες ψωμιού και ως Negative ορίζονται οι εικόνες λαχανικών. Το μοντέλο κλήθηκε να κατηγοριοποιήσει 50 παραγόμενες εικόνες ψωμιού και 50 παραγόμενες εικόνες λαχανικών (test dataset).

Στους παρακάτω πίνακες (πίνακες 4.11 και 4.12) υπολογίζονται οι τιμές για τα True positive rate, False negative rate, False positive rate και το True negative rate αλλά και για το Accuracy, το Balanced accuracy, το Positive Predictive value, Negative predict value και το F1 score ώστε να αποφανθούμε κατά πόσο το μοντέλο μπόρεσε να κατηγοριοποιήσει σωστά τις παραγόμενες εικόνες.

Training epochs	True positive rate (TPR)	False negative rate (FNR)	False positive rate (FPR)	True negative rate (TNR)
20	1	0	0.26	0.74

Πίνακας 4. 11: Παρουσιάζονται οι τιμές για το True positive rate, False negative rate, False positive rate και το True negative rate για την εποχή 20. Ως Positive ορίζονται οι εικόνες ψωμιού και ως Negative ορίζονται οι εικόνες λαχανικών. Το μοντέλο κλήθηκε να κατηγοριοποιήσει 50 παραγόμενες εικόνες ψωμιού και 50 παραγόμενες εικόνες λαχανικών (test dataset). Έγινε στρογγυλοποίηση δύο δεκαδικών ψηφίων.

Training epochs	Accuracy (ACC)	Balanced accuracy (BA)	Positive Predictive value (PPV)	Negative predict value (NPV)	F1 score
20	0.87	0.87	0.79	1	0.88

Πίνακας 4. 12: Παρουσιάζονται οι τιμές για το Accuracy, το Balanced accuracy, το Positive Predictive value, Negative predict value και το F1 score για την εποχή 20. Ως Positive ορίζονται οι εικόνες ψωμιού και ως Negative ορίζονται οι εικόνες λαχανικών. Το μοντέλο κλήθηκε να κατηγοριοποιήσει 50 παραγόμενες εικόνες ψωμιού και 50 παραγόμενες εικόνες λαχανικών (test dataset). Έγινε στρογγυλοποίηση δύο δεκαδικών ψηφίων.

Καταλήγουμε λοιπόν ότι το μοντέλο κατάφερε να αναγνωρίσει με αρκετά μεγάλη επιτυχία τις παραγόμενες εικόνες ψωμιού και λαχανικών του GAN αυτής της εργασίας. Συγκεκριμένα η ακρίβεια (Accuracy (ACC)) στην ταξινόμηση των εικόνων έφτασε το 87%. Η τιμή του Positive Predictive value (PPV) έφτασε το 0.79 δηλαδή κατά 79% εάν το μοντέλο προβλέψει ότι μια εικόνα είναι εικόνα ψωμιού, η συγκεκριμένη εικόνα είναι πράγματι εικόνα ψωμιού. Το Negative predict value (NPV) πήρε τιμή 1 αφού αν το μοντέλο αναγνωρίζει μια εικόνα ως εικόνα λαχανικών σίγουρα η εικόνα αυτή δεν θα είναι εικόνα ψωμιού. Τέλος και το F1 score παίρνει τιμή 0.88.

Συγκρίνοντας τις τιμές αυτές με τις αντίστοιχες τιμές για το πείραμα που εκτελέστηκε στις ίδιες εποχές και παραμέτρους, αλλά αξιολογήθηκε με το validation dataset, παρατηρούμε ορισμένες διαφορές. Συγκεκριμένα το μοντέλο που αξιολογήθηκε με το validation dataset κατάφερε να κατηγοριοποιήσει όλες τις πραγματικές εικόνες με επιτυχία έχοντας Accuracy ίσο με 1. Από την άλλη πλευρά όταν το μοντέλο κλήθηκε να ταξινομήσει τις παραγόμενες εικόνες του GAN τα κατάφερε με 87% επιτυχία. Αυτό οφείλεται στο ότι οι παραγόμενες εικόνες, όπως είναι λογικό, είναι χαμηλότερης ποιότητας από της πραγματικές, πράγμα που δυσκολεύει το CNN μοντέλο να επιτύχει το ίδιο σκορ στην ταξινόμηση των παραγόμενων εικόνων. Ωστόσο καταφέρνει να τις κατηγοριοποιήσει σε ένα αρκετά μεγάλο ποσοστό (87%) κάτι που σημαίνει ότι το GAN δημιούργησε εικόνες αρκετά καλές που ένα CNN μοντέλο εκπαιδευμένο σε πραγματικές εικόνες ψωμιού και λαχανικών μπόρεσε να τις αναγνωρίσει. Φυσικά επειδή το επιθυμητό θα ήταν να έχει την ικανότητα να αναγνωρίζει όλες τις παραγόμενες εικόνες με την ίδια ευκολία όπως και τις πραγματικές, το GAN φαίνεται να έχει περιθώρια βελτίωσης.

4.3 Σύγκριση αποτελεσμάτων για τα δύο πειράματα.

Παρατηρώντας τα αποτελέσματα των δύο πειραμάτων, δηλαδή τις τιμές στις μετρικές όταν το CNN μοντέλο έχει εκπαιδευτεί και αξιολογηθεί στις κλάσεις ψωμιού και μακαρονιών και την δεύτερη φορά στις κλάσεις ψωμιού και λαχανικών, λαμβάνουμε ορισμένα συμπεράσματα. Αρχικά παρατηρώντας τα διαγράμματα 4.1, 4.2, 4.3 και 4.4, 4.5, 4.6 βλέπουμε ότι και στα δύο πειράματα η εκπαίδευση τους πραγματοποιείται καλύτερα με τις ίδιες τιμές για τις παραμέτρους learning rate (learning rate = 0.0001) και batch size (batch size = 32). Έπειτα είναι φανερό ότι στο πρώτο πείραμα (διάγραμμα 4.1) χρειάζονται πολλές παραπάνω εποχές ώστε να αρχίσει να αποκτά χαμηλές τιμές το “loss” και υψηλές το “accuracy”, συγκεκριμένα από την 100ή εποχή και έπειτα. Αντίθετα στο δεύτερο πείραμα οι εποχές που χρειάζεται για να εκπαιδευτεί είναι πολύ λιγότερες, συγκεκριμένα από την 10η εποχή αρχίζει να έχει υψηλές τιμές στο accuracy. Επίσης στο πείραμα 1 υπάρχουν πολλές διακυμάνσεις στις τιμές για τα διαγράμματα της εκπαίδευσης και της αξιολόγησης της εκπαίδευσης σε σχέση με τα διαγράμματα για το πείραμα 2.

Έχοντας λοιπόν βρει σε ποια εποχή ολοκληρώνεται η καλύτερη πιθανή εκπαίδευση του μοντέλου παρατηρούμε ορισμένα πράγματα. Πρώτα απ’ όλα στο πείραμα 1 χρειάστηκαν 150 εποχές ώστε να ολοκληρωθεί η εκπαίδευσή του ενώ στο πείραμα 2 μόλις μόνο 20 εποχές. Έπειτα αξίζει να αναφερθεί ότι το μοντέλο παρουσίασε υψηλή απόδοση για το δεύτερο πείραμα, καταφέροντας να αναγνωρίσει 100% των πραγματικών εικόνων ψωμιού και λαχανικών. Φυσικά κατά την διαδικασία στην οποία το μοντέλο κατηγοριοποιούσε πραγματικές εικόνες. Από την άλλη πλευρά, στο πείραμα 2 το μοντέλο δεν παρουσίασε

πολύ υψηλή απόδοση ακόμα και όταν η εκπαίδευσή του έγινε σε πολύ μεγάλο αριθμό εποχών (150 εποχές) συγκριτικά με τις 20 εποχές που ολοκληρώθηκε η εκπαίδευση του πειράματος 1.

Στην συνέχεια δοκιμάσαμε να βάλουμε ως είσοδο στο ήδη εκπαιδευμένο μοντέλο τις παραγόμενες εικόνες του GAN ώστε να παρατηρήσουμε τα επίπεδα αναγνώρισης που θα κάνει για κάθε ένα πείραμα. Αρχικά και στα δύο πειράματα το μοντέλο φαίνεται πως αναγνώρισε τις παραγόμενες εικόνες του GAN με ποσοστό μεγαλύτερο του 50%. Ειδικότερα στο πρώτο πείραμα που κλήθηκε να ταξινομήσει τις παραγόμενες εικόνες ψωμιού και μακαρονιών κατηγοριοποίησε το 71% των εικόνων σωστά. Ακόμα μεγαλύτερο σκορ σημειώθηκε στο δεύτερο πείραμα όπου το 87% των παραγόμενων εικόνων ψωμιού και λαχανικών αναγνωρίστηκαν με επιτυχία.

Οι διαφορές που αναφέρονται παραπάνω οφείλονται όπως έχει ήδη αναφερθεί στο γεγονός ότι οι εικόνες ψωμιού διαφέρουν πολύ περισσότερο από τις εικόνες των λαχανικών συγκριτικά με τις εικόνες μακαρονιών. Αυτό ισχύει και στις πραγματικές εικόνες όπου και επιβεβαιώνεται από το γεγονός ότι το μοντέλο έκανε 130 περισσότερες εποχές ώστε να καταφέρει να εκπαιδευτεί και ξεχωρίζει τις εικόνες ψωμιού και μακαρονιών σε σχέση με τις εικόνες ψωμιού και λαχανικών. Αντίστοιχα το ότι το ποσοστό κατά το οποίο πετυχαίνει σωστή ταξινόμηση των εικόνων είναι μεγαλύτερο στο πείραμα 2 φανερώνει ότι και στις παραγόμενες εικόνες οι διαφορές ανάμεσα στις παραγόμενες εικόνες λαχανικών και ψωμιού είναι περισσότερες. Ένας από τους λόγους που υπάρχουν διαφορές στα δύο πειράματα οφείλεται στο ότι οι εικόνες του ψωμιού συγκριτικά με εκείνες των λαχανικών έχουν πολλά διαφορετικά στοιχεία με κυριότερο το χρώμα, κάτι που καθιστά την κατηγοριοποίησή τους ευκολότερη συγκριτικά με την ταξινόμηση των εικόνων ψωμιού – μακαρονιών.

Ωστόσο βλέποντας τα υψηλά σκορ που συγκέντρωσε το μοντέλο ταξινομώντας τις παραγόμενες εικόνες ψωμιού και μακαρονιών αλλά και ψωμιού και λαχανικών, γίνεται κατανοητό ότι το GAN κατάφερε να δημιουργήσει αρκετά καλές εικόνες ώστε ένα μοντέλο CNN εκπαιδευόμενο με τις κατάλληλες προδιαγραφές να καταφέρνει να τις αναγνωρίσει.

Κεφάλαιο 5: Συμπεράσματα και Μελλοντικές Προεκτάσεις της Εργασίας

Σε αυτή την εργασία σκοπός μας ήταν να δημιουργήσουμε το “έξυπνο ψυγείο”. Αρχικά ο χρήστης έχει την δυνατότητα να τραβήξει φωτογραφία το ράφι του ψυγείου του και έπειτα η φωτογραφία αυτή να αναλυθεί από το Google Vision API. Η πλατφόρμα αυτή της Google επιστρέφει σε μορφή κειμένου-ετικετών, τα τρόφιμα που περιλαμβάνει η φωτογραφία και που αναγνωρίζει. Με δεδομένο αυτές τις ετικέτες και αφού επιλέξουμε τις επιθυμητές εκπαιδεύουμε ένα Generative Adversarial Network με το αντίστοιχο σύνολο δεδομένων κάθε φορά. Καταφέρνουμε λοιπόν να δημιουργήσουμε νέες συνθετικές εικόνες οι οποίες αναπαριστούν το τι θα μπορούσε να μαγειρέψει ο χρήστης αν χρησιμοποιούσε τα συγκεκριμένα συστατικά που αναγνωρίστηκαν. Φαίνεται λοιπόν ότι καταφέραμε να δημιουργήσουμε εικόνες που πλησιάζουν οπτικά αντίστοιχες πραγματικές εικόνες του αποτελέσματος που περιμέναμε. Επιπλέον εκπαιδεύσαμε ένα Convolutional Neural Network σε δυο κλάσεις και έπειτα δώσαμε σαν είσοδο εικόνες συνθετικές ώστε να αξιολογηθεί η απόδοση του GAN μας. Φάνηκε ότι το μοντέλο κατάφερε να ταξινομήσει αρκετά καλά τις παραγόμενες εικόνες. Αυτό αποδεικνύει ότι το GAN δημιούργησε αξιόλογες εικόνες που εκτός από το ανθρώπινο μάτι, μπορούν εύκολα να κατηγοριοποιηθούν και από ένα νευρωνικό μοντέλο.

Μελλοντικά, μπορεί να βελτιωθεί η ιδέα αυτή. Αρχικά μια καλή αρχή θα ήταν να βελτιωθεί το GAN ώστε να παράγει εικόνες με μεγαλύτερη ανάλυση. Στην παρούσα διπλωματική καταφέραμε να δημιουργούμε εικόνες 128x128 pixels. Η δημιουργία εικόνων με περισσότερα pixels θα βελτίωνε την συνθετική εικόνα αφού θα απεικόνιζε περισσότερες λεπτομέρειες. Φυσικά για να επιτευχθεί αυτό απαιτούνται επιπλέον υπολογιστικοί πόροι και συγκεκριμένα είναι απαραίτητη η χρήση τουλάχιστον δύο GPUs. Επιπλέον καλό θα ήταν το μοντέλο αυτό να εκπαιδεύεται με μεγαλύτερα σύνολα δεδομένων για κάθε κατηγορία φαγητού. Εμείς εισάγουμε στο δίκτυο 200 έγχρωμες εικόνες για κάθε κατηγορία φαγητού. Ο αριθμός αυτός είναι αρκετά μικρός για εκπαίδευση Generative Adversarial Networks και γι’ αυτό θα θέλαμε ιδανικά αυξηθεί κατά πολύ ο αριθμός του συνόλου δεδομένων εκπαίδευσης για κάθε κατηγορία.

Βιβλιογραφία

- [1] S. Li, L. Da Xu, and S. Zhao, “The internet of things: a survey,” *Inf. Syst. Front.*, vol. 17, no. 2, pp. 243–259, 2015, doi: 10.1007/s10796-014-9492-7.
- [2] C. Harrold, “Practical Smart Device Design and Construction,” *Pract. Smart Device Des. Constr.*, pp. 3–14, 2020, doi: 10.1007/978-1-4842-5614-5.
- [3] K. Zarkogianni and K. S. Nikita, “Special issue on emerging technologies for the management of diabetes mellitus,” *Med. Biol. Eng. Comput.*, vol. 53, no. 12, 2015.
- [4] K. Zarkogianni *et al.*, “Comparative assessment of glucose prediction models for patients with type 1 diabetes mellitus applying sensors for glucose and physical activity monitoring,” *Med. Biol. Eng. Comput.*, vol. 53, no. 12, pp. 1333–1343, 2015, doi: 10.1007/s11517-015-1320-9.
- [5] K. Zarkogianni *et al.*, “A Review of Emerging Technologies for the Management of Diabetes Mellitus,” *IEEE Trans. Biomed. Eng.*, vol. 62, no. 12, pp. 2735–2749, 2015, doi: 10.1109/TBME.2015.2470521.
- [6] S. Luo, J. S. Jin, and J. Li, “A smart fridge with an ability to enhance health and enable better nutrition,” *Int. J. Multimed. Ubiquitous Eng.*, vol. 4, no. 2, pp. 69–80, 2009.
- [7] K. Mitsis, K. Zarkogianni, N. Bountouni, M. Athanasiou, and K. S. Nikita, “An Ontology-Based Serious Game Design for the Development of Nutrition and Food Literacy Skills,” *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBS*, pp. 1405–1408, 2019, doi: 10.1109/EMBC.2019.8856604.
- [8] K. Mitsis, K. Zarkogianni, K. Dalakleidi, G. Mourkousis, and K. S. Nikita, “Evaluation of a serious game promoting nutrition and food literacy: Experiment design and preliminary results,” *Proc. - 2019 IEEE 19th Int. Conf. Bioinforma. Bioeng. BIBE 2019*, pp. 497–502, 2019, doi: 10.1109/BIBE.2019.00096.
- [9] T. Salvador, Χάινς, Biswas, Αϊτάρ, Μαρίν, Όφλι, Βέμπερ, “Recipe1M+: A Dataset for Learning Cross-Modal Embeddings for Cooking Recipes and Food Images.” <http://pic2recipe.csail.mit.edu/> (accessed Jun. 18, 2022).
- [10] Fangda Han, Ricardo Guerrero, Vladimir Pavlovic, “CookGAN: Meal image synthesis from ingredients,” *Proc. - 2020 IEEE Winter Conf. Appl. Comput. Vision, WACV 2020*, pp. 1439–1447, 2020, doi: 10.1109/WACV45572.2020.9093463.
- [11] M. Sabini, “GAN-stromony : Generative Cooking with Conditional DCGANs”.
- [12] Q. Y. Hu, C. D. Wang, J. X. Hong, M. Z. Hua, and D. Huang, *Collaborative Computing: Networking, Applications, and Worksharing*, vol. 163. 2016. [Online]. Available: <http://www.scopus.com/inward/record.url?eid=2-s2.0-84958050992&partnerID=tZOtx3y1>
- [13] Δ. Καράνι, “Introduction to Word Embedding and Word2Vec.” <https://towardsdatascience.com/introduction-to-word-embedding-and-word2vec-652d0c2060fa> (accessed Jun. 18, 2022).
- [14] Wikipedia, “Long short-term memory.” https://en.wikipedia.org/wiki/Long_short-term_memory
- [15] Wikipedia, “Gated recurrent unit.” https://en.wikipedia.org/wiki/Gated_recurrent_unit (accessed Jun. 18, 2022).
- [16] Y. Ito, W. Shimoda, and K. Yanai, “Food image generation using a large amount of food images with conditional GAN: RamenGAN and RecipeGAN,” *ACM Int. Conf. Proceeding Ser.*, no. 1, pp. 71–74, 2018, doi: 10.1145/3230519.3230598.
- [17] D. Karagiannis and K. S. Nikita, “Design and development of a 3D Printed IoT portable Pillbox for continuous medication adherence,” *Proc. - 2020 IEEE Int. Conf. Smart Internet Things, SmartIoT 2020*, pp. 352–353, 2020, doi: 10.1109/SmartIoT49966.2020.00066.
- [18] S. Madakam, “Internet of Things: Smart Things,” *Int. J. Futur. Comput. Commun.*, vol. 4, no. 4, pp. 250–253, 2015, doi: 10.7763/ijfcc.2015.v4.395.
- [19] A. Kevin, “That ’ Internet of Things ’ Thing,” *RFID J.*, p. 4986, 2010, [Online]. Available: <http://www.itco.jp/libraries/RFIDjournal-That Internet of Things Thing.pdf>
- [20] D. Karagiannis, I. Maglogiannis, K. S. Nikita, and P. Tsanakas, “Hardware/Software Co-Design of a Low-Power IoT Fall Detection Device,” *IFIP International Internet of Things Conference, 2021*, pp. 146–159.
- [21] W. D. Fang, W. He, W. Chen, L. H. Shan, and F. Y. Ma, “Research on the application-driven architecture in internet of things,” *Front. Artif. Intell. Appl.*, vol. 293, pp. 458–465, 2016, doi: 10.3233/978-1-61499-722-1-458.
- [22] Wikipedia, “IPv6.” <https://el.wikipedia.org/wiki/IPv6> (accessed Jun. 18, 2022).
- [23] A. Montazerolghaem, M. H. Yaghmaee, and A. Leon-Garcia, “Green Cloud Multimedia Networking: NFV/SDN Based Energy-Efficient Resource Allocation,” *IEEE Trans. Green Commun. Netw.*, vol. 4, no. 3, pp. 873–889, 2020, doi: 10.1109/TGCN.2020.2982821.
- [24] I. C. Education, “Application Programming Interface (API).” <https://www.ibm.com/cloud/learn/api> (accessed Jun. 26, 2022).
- [25] M. E. P. Hernández and S. Reiff-Marganiec, “Classifying smart objects using capabilities,” *Proc. 2014 Int. Conf. Smart Comput. SMARTCOMP 2014*, pp. 309–316, 2014, doi: 10.1109/SMARTCOMP.2014.7043873.
- [26] J. L. M. R. I Hartley, “The relationship between photogrammetry and computer vision,” vol. 1944, pp. 92–105, 1944.
- [27] A. N. Ramesh, C. Kambhampati, J. R. T. Monson, and P. J. Drew, “Artificial intelligence in medicine,” *Ann. R. Coll. Surg. Engl.*, vol. 86, no. 5, pp. 334–338, 2004, doi: 10.1308/147870804290.

- [28] A. Entwistle, “What Is Artificial Intelligence?,” *Eng. Mater. Des.*, vol. 32, no. 3, pp. 1–14, 1988, doi: 10.7551/mitpress/12518.003.0004.
- [29] N. World, “1 . WHAT IS ARTIFICIAL INTELLIGENCE ?,” no. 1988, pp. 3–27, 1990.
- [30] I. El Naqa and M. J. Murphy, “Machine Learning in Radiation Oncology,” *Mach. Learn. Radiat. Oncol.*, pp. 3–11, 2015, doi: 10.1007/978-3-319-18305-3.
- [31] Α. ΧΡΟΝΟΠΟΥΛΟΥ, “Τεχνικές Μεταφοράς Μάθησης σε Βαθιά Νευρωνικά Δίκτυα για Ανάλυση Συναισθήματος και Σημσιολογική Μοντελοποίηση,” 2019. [Online]. Available: http://artemis.cslab.ece.ntua.gr:8080/jspui/bitstream/123456789/17239/1/diplomatiki_achronopoulou.pdf
- [32] M. I. Jordan and T. M. Mitchell, “Machine learning: Trends,perspectives, and prospects,” *Science (80-.)*, vol. 349, no. 6245, pp. 255–260, 2015.
- [33] P. D. McNicholas and P. A. Tait, “Supervised Learning,” *Data Sci. with Julia*, pp. 93–128, 2019, doi: 10.1201/9781351013673-5.
- [34] B. Liu, *Web Data Mining*. 2011. doi: 10.1007/978-3-642-19460-3.
- [35] V. Verdhan, “Introduction to Supervised Learning,” *Supervised Learn. with Python*, pp. 1–46, 2020, doi: 10.1007/978-1-4842-6156-9_1.
- [36] T. Berg-Kirkpatrick, A. Bouchard-Côté, J. De Nero, and D. Klein, “Painless unsupervised learning with features,” *NAACL HLT 2010 - Hum. Lang. Technol. 2010 Annu. Conf. North Am. Chapter Assoc. Comput. Linguist. Proc. Main Conf.*, no. June, pp. 582–590, 2010.
- [37] Νικόλ Ρασκ, “Deep learning.” <https://www.nature.com/articles/nmeth.3707> (accessed Jun. 18, 2022).
- [38] Α. Πλέρου, “Τεχνητά Νευρωνικά Δίκτυα Προσομοίωσης Του Ανθρώπινου Εγκεφάλου,” *Ανοικτή Εκπαίδευση Το Περιοδικό Για Την Ανοικτή Και Εξ Αποστάσεως Εκπαίδευση Και Την Εκπαιδευτική Τεχνολογία*, vol. 8, no. 1, p. 128, 2012, doi: 10.12681/jode.9794.
- [39] A. Karpathy, “Convolutional neural networks for visual recognition.” <https://cs231n.github.io/neural-networks-1/> (accessed Jun. 18, 2022).
- [40] F.-X. W. Sakib Mostafa, “Chapter 3 - Diagnosis of autism spectrum disorder with convolutional autoencoder and structural MRI images,” in *Neural Engineering Techniques for Autism Spectrum Disorder*, 2021. doi: <https://doi.org/10.1016/C2019-0-05414-3>.
- [41] K. O’Shea and R. Nash, “An Introduction to Convolutional Neural Networks,” pp. 1–11, 2015, [Online]. Available: <http://arxiv.org/abs/1511.08458>
- [42] A. Papadopoulos and University of Patras, “Convolutional Neural Networks and their usage on computer vision,” 2016.
- [43] J. Brownlee, “A Gentle Introduction to the Rectified Linear Unit (ReLU),” 2020. <https://machinelearningmastery.com/rectified-linear-activation-function-for-deep-learning-neural-networks/> (accessed Jun. 26, 2022).
- [44] Α. Γεώργιος, “Σύνθεση εικόνων από Sky Camera Images με βαθιά νευρωνικά δίκτυα τύπου GAN (Generative Adversarial Networks).”
- [45] S. Barua, S. M. Erfani, and J. Bailey, “FCC-GAN: A Fully Connected and Convolutional Net Architecture for GANs,” no. ii, 2019, [Online]. Available: <http://arxiv.org/abs/1905.02417>
- [46] M. Mirza and S. Osindero, “Conditional Generative Adversarial Nets,” pp. 1–7, 2014, [Online]. Available: <http://arxiv.org/abs/1411.1784>
- [47] Φ. Κονιδάρης and Σ. Ανδρέας-Γεώργιος, “Χρήση Γεννητικών Ανταγωνιστικών Δικτύων ως Τεχνική Εμπλουτισμού Δεδομένων σε Ιατρικές Εφαρμογές,” Εθνικό Μετσόβιο Πολυτεχνείο, 2018. [Online]. Available: <http://artemis.cslab.ece.ntua.gr:8080/jspui/handle/123456789/17038>
- [48] J. D. Katsilometes, “How good is my maintenance program?,” *Improv. Optim. Oper. Things That Actually Work - Plant Oper. Forum 2004*, no. 5302, pp. 27–37, 2005.
- [49] A. Kulkarni, D. Chong, and F. A. Batareseh, *Foundations of data imbalance and solutions for a data democracy*. Elsevier Inc., 2020. doi: 10.1016/B978-0-12-818366-3.00005-8.
- [50] A. S. Pushpa Singh,Narendra Singh,Krishna KantSingh, *Chapter 5 - Diagnosing of disease using machine learning*. 2021. doi: 10.1016/B978-0-12-821229-5.00003-3.
- [51] A. Radford, L. Metz, and S. Chintala, “Unsupervised representation learning with deep convolutional generative adversarial networks,” *4th Int. Conf. Learn. Represent. ICLR 2016 - Conf. Track Proc.*, pp. 1–16, 2016.
- [52] S. Swanrkar, “Image-Generator.” <https://github.com/shivamswanrkar/Image-Generator> (accessed Jun. 18, 2022).
- [53] K. Kogias, I. Andreadis, K. Dalakleidi, and K. S. Nikita, “A Two-Level Food Classification System for People with Diabetes Mellitus Using Convolutional Neural Networks,” *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBS*, vol. 2018-July, pp. 2603–2606, 2018, doi: 10.1109/EMBC.2018.8512839.

