

ΤΜΗΜΑΤΟΠΟΙΗΣΗ ΟΔΟΝΤΙΚΩΝ ΕΙΚΟΝΩΝ ΜΕ
ΤΗΝ ΧΡΗΣΗ ΝΕΥΡΩΝΙΚΩΝ ΔΙΚΤΥΩΝ

Χρήστος Φιλιππής

Πανεπιστήμιο Πειραιώς
Τμήμα Ψηφιακών Συστημάτων
Μεγάλα Δεδομένα και Αναλυτική

Επιβλέπων : Ηλίας Μαγκλογιάννης

Ιούνιος 2020

1 Περιεχόμενα

2	Εισαγωγή.....	4
2.1	Αντικείμενο της Μεταπτυχιακής Διατριβής.....	4
2.2	Σκοπός της διατριβής.....	5
2.3	Δομή της διατριβής.....	5
3	Τεχνικό Υπόβαθρο και Επισκόπηση Βιβλιογραφίας.....	6
3.1	Μηχανική μάθηση.....	6
3.2	Είδη Μηχανικής Μάθησης.....	7
3.2.1	Εκπαίδευση αλγορίθμου πάνω σε εικόνα.....	8
3.3	CNN.....	9
3.4	Mask RCNN.....	10
3.5	Μετρικές αξιολόγησης.....	14
3.5.1	Mean Average Precision.....	14
3.5.2	Ακρίβεια και ανάκληση.....	14
3.5.3	IoU (Intersection over union).....	14
3.6	Γλώσσα και βιβλιοθήκες προγραμματισμού.....	15
4	Προτεινόμενη Υλοποίηση.....	16
4.1	Υποδομή και εργαλεία.....	16
4.1.1	Υποδομή.....	16
4.1.2	Docker.....	16
4.2	Υλοποίηση.....	16
4.2.1	Σύνοψη υλοποίησης.....	16
4.2.2	Δομή και αρχεία προγράμματος.....	17
4.2.3	Αρχικοποίηση.....	19
4.2.4	Προεπεξεργασία δεδομένων.....	20
4.2.5	Παραμετροποίηση εκπαίδευσης.....	21
4.2.6	Πρόγραμμα εκπαίδευσης.....	22
4.2.7	Αύξηση δεδομένων (Data Augmentation).....	23
4.2.8	Συνδυασμός και προσαρμογή.....	23
4.2.9	Αξιολόγηση.....	24
5	Πειραματική Αξιολόγηση και Αποτελέσματα.....	26
5.1	Δεδομένα.....	26
5.1.1	Διαδικασία λήψης εικόνας.....	26
5.1.2	Εργαλείο χαρακτηρισμού εικόνων.....	27
5.2	Μετρικές απόδοσης Mask RCNN.....	29
5.3	Αποτελέσματα.....	29
5.3.1	Δοκιμές στο πρόγραμμα εκπαίδευσης.....	30

5.3.2	Δοκιμές στην προσαρμογή των προβλέψεων	38
5.3.3	Δοκιμές σε προσαρμοσμένες κλάσεις	43
6	Συμπεράσματα.....	47
7	Ευχαριστίες.....	48
8	Παραρτήματα	49
8.1	Παράρτημα I. Dockerfile.....	49
8.2	Παράρτημα II. Εντολές εκκίνησης εφαρμογής	49
8.3	Παράρτημα III. Πρόγραμμα εκπαίδευσης.....	50
8.4	Παράρτημα IV. Βοηθητικές βιβλιοθήκες προγραμματισμού σε Python.....	50
9	Αναφορές.....	51

2 Εισαγωγή

2.1 Αντικείμενο της Μεταπτυχιακής Διατριβής

Το αντικείμενο της διατριβής αυτής είναι να βρεθεί ένας τρόπος προκειμένου να γίνονται ακριβείς διαγνώσεις πάνω σε δόντια χωρίς να είναι επιβλαβείς για τον ασθενή. Μέχρι τώρα για να μπορέσει να γίνει μια διάγνωση για την κατάσταση ενός δοντιού το πρώτο πράγμα που γίνεται, είναι να αξιολογηθεί οπτικά από τον οδοντίατρο. Στην συνέχεια χρησιμοποιούνται ειδικά μηχανήματα προκειμένου να γίνουν κάποιες απεικονίσεις με την χρήση ακτινοβολίας.

Οι εξετάσεις τέτοιου είδους έχουν κάποια πολύ σημαντικά μειονεκτήματα. Καταρχάς για να πραγματοποιηθεί μια τέτοια εξέταση γίνεται χρήση ακτινών (συνήθως X) οι οποίες είναι επιβλαβείς για την υγεία του εξεταζόμενου. Εκτός από αυτό υπάρχουν και περιπτώσεις που για διάφορους ιατρικούς λόγους ο εξεταζόμενος δεν μπορεί εκτεθεί σε τέτοιου είδους ακτινοβολία με αποτέλεσμα να μην μπορεί να πραγματοποιήσει τις συγκεκριμένες εξετάσεις.

Εκτός από το ότι είναι επιβλαβείς οι εξετάσεις αυτές, όπως και οι περισσότερες ιατρικές εξετάσεις, προϋποθέτουν ειδικά μηχανήματα και εξοπλισμό καθώς και ειδικούς οι οποίοι μπορούν να πραγματοποιούν τις εξετάσεις αυτές με ασφάλεια. Τα παραπάνω μηχανήματα καθώς και ο εξοπλισμός έχουν πολύ μεγάλο κόστος και συνήθως τα μηχανήματα είναι μεγάλα και για να λειτουργήσουν σωστά πρέπει να βρίσκονται σε ειδικούς χώρους. Τέλος για όλα τα παραπάνω πολλές τέτοιες εξετάσεις γίνονται σε ειδικά εργαστήρια πράγμα που σημαίνει ότι τα αποτελέσματα βγαίνουν μετά από λίγες μέρες.

Οπότε η αρχική σκέψη ήταν να γίνει χρήση της μηχανικής μάθησης προκειμένου να γίνουν μη επιβλαβείς διαγνώσεις, γρήγορα, με την χρήση πολύ φθηνού εξοπλισμού. Οπότε βασική ιδέα της διατριβής είναι να φωτογραφιστεί ένα δόντι με μια μικρή κάμερα, στην συνέχεια με την χρήση ενός υπολογιστή (ακόμα και ενός έξυπνου τηλεφώνου) να αναλυθεί η εικόνα και να βγάλει κάποια αποτελέσματα για πιθανές βλάβες στο δόντι που φωτογραφίστηκε.

Με τον τρόπο αυτό υπάρχουν θα μπορούσαν να υπάρχουν ανάλογα αποτελέσματα με μια εξέταση από ένα μηχάνημα που χρησιμοποιεί ακτίνες για τις απεικονίσεις του, όμως τα αποτελέσματα θα βγαίνουν αμέσως και ο εξοπλισμός που θα χρειαζόταν είτε βρίσκεται ήδη σε κάθε σπίτι (ηλεκτρονικός υπολογιστής), είτε θα μπορούσε να έχει ένα μικρό κόστος, σε αντίθεση με τα ακριβά μηχανήματα απεικονίσεων. Τέλος και ίσως σημαντικότερο δεν θα επιβάρυνε την υγεία του εξεταζόμενου με ακτινοβολία.

Η παραπάνω διαδικασία δεδομένου ότι είναι πολύ εύκολη δεν χρειάζεται την επέμβαση κάποιου ειδικού είτε γιατρού, είτε κάποιου ακτινολόγου και μπορεί να εκτελεστεί ενδεχομένως και από τον ίδιο τον εξεταζόμενο. Αυτό είναι πολύ σημαντικό γιατί μπορεί να διευκολύνει τις ιατρικές διαγνώσεις σε απομονωμένες περιοχές, όπως μικρά νησιά ή περιοχές που αποκλείονται λόγω έντονων καιρικών φαινομένων (χιονοπτώσεις, έντονα φαινόμενα στις θάλασσα), και να προληφθούν διαφόρων καταστάσεων.

Οπότε συνοψίζοντας στην διατριβή αυτή γίνεται μια προσπάθεια για διαγνώσεις που δεν είναι επιβλαβείς, γίνονται εύκολα και με μικρό κόστος και βγαίνουν αμέσως τα αποτελέσματα. Οι εξετάσεις αυτές δεν χρειάζονται την παρουσία κάποιου ειδικού οπότε ευνοούν και την τηλεϊατρική.

2.2 Σκοπός της διατριβής

Ο σκοπός της διατριβής αυτής είναι να γίνει χρήση της μηχανικής μάθησης προκειμένου να εξάγουμε πολύτιμες πληροφορίες από μια εικόνα και στην συνέχεια να χρησιμοποιηθούν για ιατρικούς σκοπούς. Οπότε εκπαιδεύεται ένα μοντέλο το οποίο θα εντοπίζει περιοχές ενδιαφέροντος σε εικόνες από δόντια. Οι περιοχές ενδιαφέροντος είναι βλάβες σε δόντια και κατά κύριο λόγο επιφανειακές και αφορούν παθήσεις των δοντιών όπως η τερηδόνα.

Σαν βάση χρησιμοποιείται ένα ισχυρό νευρωνικό δίκτυο και στην συνέχεια γίνεται μια προσπάθεια βελτιστοποίησης των αποτελεσμάτων με περαιτέρω επεξεργασία των προβλέψεων που προέκυψαν από το νευρωνικό. Το νευρωνικό που επιλέχθηκε έχει πολύ καλή απόδοση σε γνωστά σύνολα δεδομένων όπως το COCO [1] με μεγάλο πλήθος αντικειμένων και στιγμιότυπων. Οπότε σαν αρχικό μοντέλο χρησιμοποιείται ένα μοντέλο που έχει προκύψει από εκπαίδευση πάνω σε αυτό το σύνολο δεδομένων, και στην συνέχεια προσαρμόστηκε πάνω στα δεδομένα του συγκεκριμένου προβλήματος.

Εκτός από την πολυπλοκότητα στο να εκπαιδευτεί ένας τέτοιο μοντέλο και τους πόρους που χρειάζεται για να εκπαιδευτεί, ένα άλλο πρόβλημα που αντιμετωπίζεται είναι το μικρό σύνολο δεδομένων, δεδομένου ότι είναι δύσκολο να παραχθούν μιας και χρειάζονται την γνώμη ενός ειδικού. Για να αντιμετωπιστεί το παραπάνω χρησιμοποιούνται διάφορες τεχνικές προκειμένου να μεγαλώσει το πλήθος των δεδομένων.

Οι εικόνες που αναλύονται είναι υψηλής ανάλυσης, έγχρωμες εικόνες από δόντια που είτε έχουν βλάβες είτε όχι. Οι εικόνες αυτές έχουν τραβηχτεί από οδοντίατρος χρησιμοποιώντας κάμερες ειδικά σχεδιασμένες για τη χρήση αυτή.

2.3 Δομή της διατριβής

Στην διατριβή αυτή σε πρώτη φάση ορίζεται το πρόβλημα που καλείται να λύσει, την χρήση δηλαδή της μηχανικής μάθησης για πιο ακριβείς ιατρικές διαγνώσεις, οι οποίες ταυτόχρονα είναι και λιγότερο επιβλαβείς. Στην συνέχεια και αφού γίνει μια αναφορά στους αλγόριθμους που θα χρησιμοποιηθούν καθώς και στην γλώσσα προγραμματισμού που επιλέχθηκε, αναλύεται η υλοποίηση που αναπτύχθηκε καθώς και η όλη μεθοδολογία έτσι ώστε να προσεγγιστεί όσο καλύτερα γίνεται το πρόβλημα που ορίστηκε νωρίτερα.

Στο κεφάλαιο με την προτεινόμενη υλοποίηση σε πρώτη φάση γίνεται μια αναφορά στην υποδομή (υλικό) την οποία επιλέχθηκε να τρέξει, τα εργαλεία ποια εργαλεία τα οποία χρησιμοποιήθηκαν έτσι ώστε η εφαρμογή να τρέχει ανεξάρτητα από το λειτουργικό σύστημα που είναι εγκατεστημένο στον υπολογιστή που την φιλοξενεί, και τέλος περιγράφεται και ένα εργαλείο που χρησιμοποιήθηκε για να δημιουργηθούν τα δεδομένα εκπαίδευσης.

Στην συνέχεια του κεφαλαίου αυτού περιγράφεται η εφαρμογή που δημιουργήθηκε, τα βήματα που ακολουθεί, την παραμετροποίηση στο νευρωνικό δίκτυο και την περαιτέρω επεξεργασία των αποτελεσμάτων που έγιναν προκειμένου να βελτιωθεί η ακρίβεια τους.

Στο κεφάλαιο που ακολουθεί παρουσιάζεται ο τρόπος με τον οποίο δημιουργήθηκαν τα δεδομένα εκπαίδευσης από γιατρούς και στην συνέχεια παρουσιάζονται κάποιες δοκιμές που έγιναν με διάφορα προγράμματα εκπαίδευσης και την απόδοση του καθενός. Τέλος γίνεται μια εκτενής αναφορά στην βιβλιογραφία που χρησιμοποιήθηκε σαν βάση για την συγκεκριμένη διατριβή.

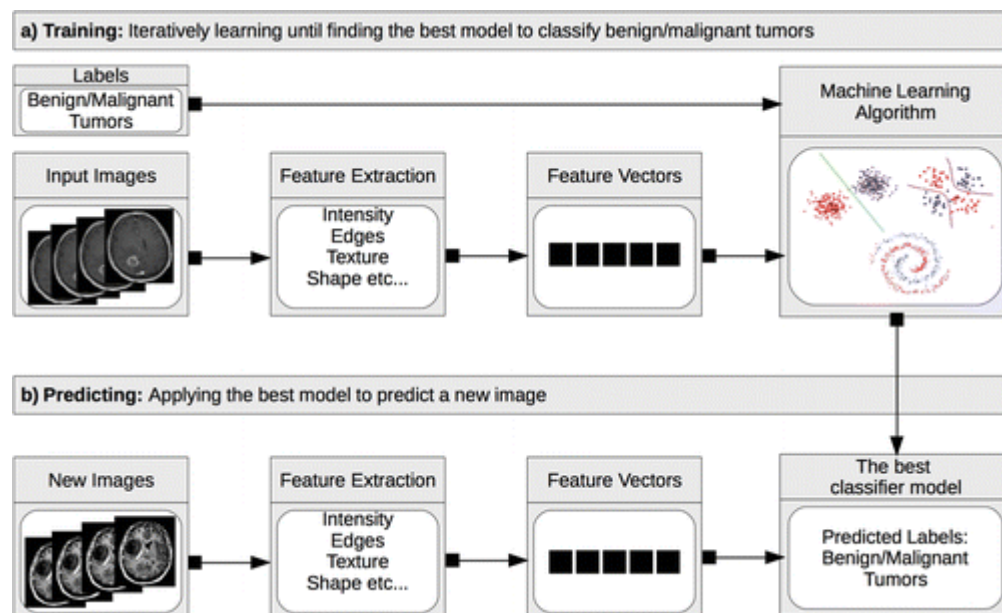
3 Τεχνικό Υπόβαθρο και Επισκόπηση Βιβλιογραφίας

3.1 Μηχανική μάθηση

Με τον όρο Μηχανική Μάθηση αναφερόμαστε στο σύνολο των μαθηματικών αλγορίθμων που χρησιμοποιούνται για την εκπαίδευση υπολογιστικών συστημάτων, τα οποία επιφορτίζονται με τη διεκπεραίωση καθημερινών εργασιών που μέχρι πρότινος εκτελούνταν συνήθως με τη χρήση ανθρώπινου έργου. Ουσιαστικά ο πυρήνας της Μηχανικής Μάθησης είναι ο αλγόριθμος εκπαίδευσης, δηλαδή ένας αλγόριθμος που έχει την ικανότητα να μάθει χρησιμοποιώντας διάφορες πηγές δεδομένων και να αυξάνει με την πάροδο του χρόνου αυτή του την ικανότητα μέσω επαναληπτικών διεργασιών.

Η Μηχανική Μάθηση αποτελεί ένα από τα επιστημονικά πεδία που απαρτίζουν το γνωστικό αντικείμενο της Τεχνητής Νοημοσύνης και μας δίνει τη δυνατότητα να αντιμετωπίσουμε πολύπλοκα προβλήματα, η επίλυση των οποίων είναι πολύ δύσκολη με τη χρήση προγραμμάτων που γράφονται και σχεδιάζονται αποκλειστικά από ανθρώπους. Από επιστημονική άποψη η Μηχανική Μάθηση παρουσιάζει πολύ μεγάλο ενδιαφέρον, καθώς η ανάπτυξη της κατανόησης μας για αυτή συνεπάγεται με την ανάπτυξη της κατανόησης μας για τις αρχές που διέπουν τη νοημοσύνη γενικότερα [2], [3].

Κάτι που δεν είναι ευρέως γνωστό είναι ότι η μηχανική μάθησης μπορεί να χρησιμοποιηθεί στην ανάλυση ιατρικών εικόνων και σε διεργασίες ερμηνείας περιοχών εικόνων [4] [5]. Το ακόλουθο είναι ένας ευρέως αποδεκτός ορισμός της μηχανικής μάθησης: Εάν ένας αλγόριθμος μηχανικής μάθησης εφαρμοστεί σε ένα σύνολο δεδομένων (στο παράδειγμα μας, εικόνες όγκων), και σε κάποια γνώση για αυτά τα δεδομένα (στο παράδειγμά μας, καλοήθεις ή κακοήθεις όγκους) τότε το σύστημα του αλγορίθμου μπορεί να μάθει από τα δεδομένα εκπαίδευσης και να εφαρμόσει ότι έχει μάθει ώστε να κάνει μια πρόβλεψη (στο παράδειγμά μας, εάν μια διαφορετική εικόνα απεικονίζει καλοήθη ή κακοήθη ιστό). Εάν ο αλγόριθμος βελτιστοποιήσει τις παραμέτρους ώστε να βελτιώσει την απόδοσή του, το οποίο και σημαίνει ότι περισσότερες περιπτώσεις δοκιμών έχουν διαγνωστεί σωστά, τότε θεωρείται ότι μαθαίνει αυτή την διεργασία.



Εικόνα 1 Μοντέλο υλοποίησης μηχανικής μάθησης και μοντέλο εφαρμογής για κατηγοριοποίηση ιατρικών εικόνων. (α) Για την εκπαίδευση, το σύστημα του αλγορίθμου μηχανικής μάθησης χρησιμοποιεί ένα σύνολο εικόνων εισόδου έτσι ώστε να αναγνωρίζει τις ιδιότητες της εικόνας, οι οποίες, αφού χρησιμοποιηθούν, θα έχουν ως αποτέλεσμα την σωστή κατηγοριοποίηση της εικόνας, το οποίο σημαίνει απεικόνιση καλοήθη ή κακοήθη όγκου, σε σύγκριση με τις δοθείσες ετικέτες για τις εικόνες εισόδου. (β) Για την πρόβλεψη, αφού το σύστημα έχει μάθει πως να κατηγοριοποιεί

τις εικόνες, το μοντέλο μάθησης εφαρμόζεται σε νέες εικόνες ώστε να συμβουλευεί ραδιολόγους στην αναγνώριση του τύπου των όγκων.

Η μηχανική μάθησης εφαρμόζεται πλέον σε πολλούς τομείς, εκτός της ιατρικής, στους οποίους και έχει κεντρικό ρόλο σε διεργασίες όπως η αναγνώριση φωνής και η μετάφραση ανάμεσα σε διαφορετικές γλώσσες, η αυτόνομη πλοήγηση οχημάτων και οι συστάσεις προϊόντων. Κάποιες από αυτές τις διεργασίες δεν ήταν εφικτές στο παρελθόν, όμως η πρόσφατη πρόοδος στην μηχανική μάθησης πλέον τις καθιστούν εφικτές.

3.2 Είδη Μηχανικής Μάθησης

Στο παρελθόν, η μηχανική μάθησης απαιτούσε δομημένη είσοδο, και κάποιες τεχνικές δεν επέτρεπαν την επιτυχή μάθηση στην περίπτωση που απουσίαζε ακόμη και ένα μικρό μέρος από τα δεδομένα. Νεότεροι αλγόριθμοι μπορούν να διαχειριστούν χωρίς πρόβλημα τις παραλείψεις στα δεδομένα, και σε κάποιες περιπτώσεις μάλιστα, το σύστημα μπορεί σκόπιμα να δημιουργεί παραλείψεις στα δεδομένα, κατά τη διάρκεια της φάσης μάθησης, για να καταστεί ο αλγόριθμος πιο ισχυρός. Οι νέοι αλγόριθμοι, σε συνδυασμό με σημαντικές αυξήσεις στην υπολογιστική ισχύ και στον όγκο των δεδομένων, έχει οδηγήσει σε ένα συνεχόμενο ενδιαφέρον όσον αφορά την μηχανική μάθησης.

Στο παράδειγμα μας, η **εποπτευόμενη μάθηση** περιλαμβάνει την δημιουργία εμπειρίας χρησιμοποιώντας εικόνες με όγκο στον εγκέφαλο, οι οποίες περιλαμβάνουν σημαντικές πληροφορίες (τύπος όγκου, καλοήθης ή κακοήθης) και εφαρμόζοντας την εμπειρία που αποκτήθηκε ώστε να προβλέψει καλοήθη και κακοήθη νεοπλασία σε καινούριες εικόνες όγκου εγκεφάλου (δεδομένα δοκιμών). Σε αυτό το παράδειγμα, στο σύστημα του αλγορίθμου θα δίνονταν πλήθος εικόνων όγκου στον εγκέφαλο, στις οποίες οι όγκοι θα χαρακτηρίζονταν είτε ως καλοήθης, είτε ως κακοήθης. Στη συνέχεια, το σύστημα θα εξετάζονταν προσπαθώντας να αναθέσει σωστά τις ετικέτες καλοήθειας ή κακοήθειας στα αποτελέσματα των νέων εικόνων. Οι εικόνες αυτές είναι το δοκιμαστικό σύνολο δεδομένων.

Παραδείγματα εποπτευόμενων αλγορίθμων μάθησης περιλαμβάνουν Support Vector Machines [6], δέντρα απόφασης [7], γραμμική παλινδρόμησης [8], λογική παλινδρόμηση [9], Naive Bayes [10] [9], k-πλησιέστερος γείτονας [11], Random Forest [12], AdaBoost και μέθοδοι με την χρήση Νευρωνικών δικτύων [13].

Με μάθηση χωρίς επιτήρηση, τα δεδομένα (π.χ. εικόνες όγκου στον εγκέφαλο) επεξεργάζονται με στόχο τον διαχωρισμό των εικόνων σε ομάδες. Για παράδειγμα, αυτές που απεικονίζουν καλοήθεις όγκους και αυτές που απεικονίζουν κακοήθεις όγκους. Η βασική διαφορά είναι πως αυτό συμβαίνει χωρίς να παρέχονται στον αλγόριθμο πληροφορίες για το τι είναι οι ομάδες. Ο αλγόριθμος ορίζει πόσες ομάδες υπάρχουν και πως να τις ξεχωρίζει.

Παραδείγματα συστημάτων αλγορίθμου χωρίς επιτήρηση περιλαμβάνουν K-means [14], Mean Shift [15], Affinity Propagation [16], Ιεραρχική συσσώρευση [17], DBSCAN (density-based spatial clustering of applications with noise) [18], Gaussian mixture modeling [19], Markov random fields [20], ISODATA (iterative self-organizing data) [21], and fuzzy C-means systems [22].

Όπως και η εποπτευόμενη μάθηση, η **ενισχυμένη μάθηση** ξεκινάει με έναν ταξινομητή ο οποίος χτίστηκε χρησιμοποιώντας δεδομένα από τις ετικέτες. Ωστόσο, στη συνέχεια δίνονται στο σύστημα δεδομένα χωρίς ετικέτα, και προσπαθεί να βελτιώσει περαιτέρω την ομαδοποίηση χαρακτηρίζοντας καλύτερα αυτά τα δεδομένα, όμοια με το πως συμπεριφέρεται και με την μάθηση χωρίς επιτήρηση. Παραδείγματα αλγορίθμων ενισχυμένης μάθησης

περιλαμβάνουν τα συστήματα Maja [23]. Σε αυτό το άρθρο, εστιάζουμε στην εποπτευόμενη μάθηση, εφόσον είναι η πιο διαδεδομένη μέθοδος εκπαίδευσης για ιατρικές εικόνες.

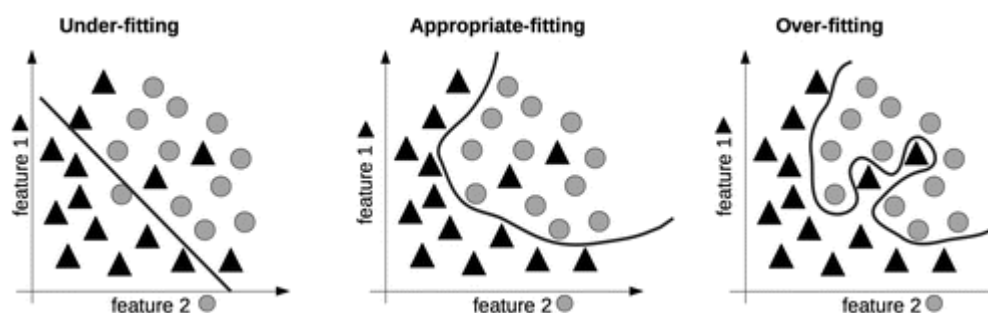
Ορισμένοι αλγόριθμοι Μηχανικής Μάθησης είναι εμπνευσμένοι από τον τρόπο λειτουργίας των βιολογικών συστημάτων μάθησης που χρησιμοποιούνται από τους έμβιους οργανισμούς, κυρίως του ανθρώπινου εγκεφάλου και τα οποία αποτελούνται από ένα πολύπλοκο δίκτυο υπερ-συνδεδεμένων νευρώνων. Οι αλγόριθμοι αυτοί ονομάζονται Τεχνητά Νευρωνικά Δίκτυα (Artificial Neural Networks) και αποτελούνται από ένα πυκνό σύνολο υπερ-συνδεδεμένων απλών μονάδων, με την κάθε μονάδα να παίρνει έναν αριθμό από εισόδους πραγματικών τιμών και αφού τις επεξεργαστεί, παράγει μία έξοδο πραγματικής τιμής.

Η διαδικασία αυτή πραγματοποιείται στα τρία επίπεδα ενός τυπικού Τεχνητού Νευρωνικού Δικτύου και τα οποία είναι το επίπεδο εισόδου, το κρυμμένο και το επίπεδο εξόδου. Υπάρχουν Τεχνητά Νευρωνικά Δίκτυα τα οποία μπορούν να έχουν περισσότερα επίπεδα, ειδικά όσον αφορά τον αριθμό των κρυμμένων επιπέδων τους, όπως και πολύ μεγάλο αριθμό νευρώνων. Η μελέτη και η δημιουργία των Τεχνητών Νευρωνικών Δικτύων που παρουσιάζουν πολύ μεγάλη πολυπλοκότητα ως προς τη δομή του συνθέτουν ένα υποείδος της Μηχανικής Μάθησης, το οποίο ονομάζεται **Βαθιά Μάθηση (Deep Learning)** [24].

3.2.1 Εκπαίδευση αλγορίθμου πάνω σε εικόνα

Κατάτμηση (Segmentation): Ο διαχωρισμός της εικόνας σε τμήματα. Για παράδειγμα, η κατάτμηση όγκου είναι η διαδικασία που ορίζει που ξεκινά και που τελειώνει ο όγκος. Ωστόσο, αυτό δεν περιλαμβάνει απαραίτητα την απόφαση πως αυτό που περιλαμβάνεται είναι όγκος. Ο στόχος σε αυτό το βήμα είναι να ορισθεί που κάτι ξεκινάει και που τελειώνει. Αυτή η τεχνική συνήθως χρησιμοποιείται με έναν ταξινομητή που καθορίζει ότι ένα τμήμα μια εικόνας απεικονίζει έναν ενισχυμένο όγκο και κάποιο άλλο τμήμα απεικονίζει έναν μη ενισχυμένο όγκο.

Overfitting: Όταν ένας ταξινομητής που είναι πολύ συγκεκριμένος στο σύνολο εκπαίδευσης δεν είναι χρήσιμος επειδή είναι γνώριμος μόνο με αυτά τα παραδείγματα, αυτό ορίζεται ως overfitting (Fig.2). Γενικά το σύνολο εκπαίδευσης πρέπει να περιλαμβάνει πολλά περισσότερα παραδείγματα πάνω από το νούμερο των συντελεστών ή μεταβλητών που χρησιμοποιούνται από τον αλγόριθμο μηχανικής μάθησης.



Εικόνα 2 Τα διαγράμματα αντιπροσωπεύουν underfitting και overfitting. Underfitting συμβαίνει όταν το μοντέλο δεν είναι πολύ απλό και δεν μπορεί να περιγράψει την ποικιλία στα δεδομένα και δεν είναι σε θέση να αναγνωρίσει κάποιο πρότυπο. Ένα ιδανικό μοντέλο αναγνωρίζει το πρότυπο αλλά δεν είναι υπερβολικά πολύ ελαστικό ή ανελαστικό για να περιγράψει τα δεδομένα. Overfitting συμβαίνει όταν το μοντέλο είναι υπερβολικά καλό τόσο ώστε να προσαρμόζεται μέχρι και σε πιθανές αποκλίσεις του συνόλου δεδομένων. Οι άξονες αντιπροσωπεύουν το χαρακτηριστικό 1 and χαρακτηριστικό 2 και αντιπροσωπεύουν τα πρώτα δύο στοιχεία του διανύσματος χαρακτηριστικών.

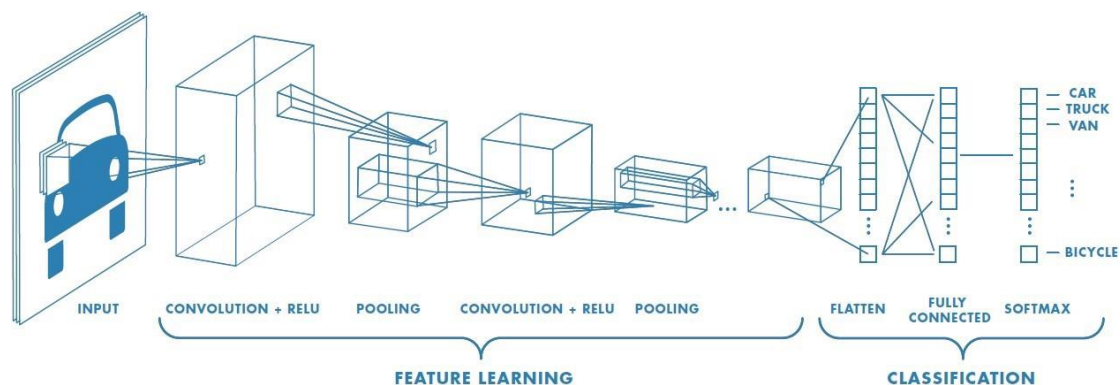
Χαρακτηριστικά (features): Αριθμητικές τιμές οι οποίες αναπαριστούν το παράδειγμα. Στην περίπτωση των ιατρικών εικόνων, τα χαρακτηριστικά μπορεί να είναι οι τιμές των εικονοστοιχείων, παραλλαγή των τιμών των εικονοστοιχείων μιας περιοχής μιας εικόνας ή

άλλες τιμές. Άλλα χαρακτηριστικά που μπορούν να χρησιμοποιηθούν είναι χαρακτηριστικά εκτός της εικόνας, όπως η ηλικία του ασθενή, ή πότε μια εργαστηριακή εξέταση είναι θετική ή αρνητική. Όταν αυτά τα χαρακτηριστικά ενός παραδείγματος συνδυαστούν, τότε λέγονται και διάνυσμα χαρακτηριστικών ή διάνυσμα εισόδου.

3.3 CNN

Το Convolutional Neural Network (CNN) είναι μία από τις δημοφιλέστερες και ευρεία χρησιμοποιημένες αρχιτεκτονικές νευρωνικών δικτύων, ειδικά σε προβλήματα που σχετίζονται με την αναγνώριση, ανάλυση και τμηματοποίηση εικόνων. Η ονομασία του προέρχεται από τη συνέλιξη (convolution) κι η οποία είναι μία μαθηματική γραμμική διεργασία μεταξύ πινάκων [25].

Παρόλο που υπάρχουν πάρα πολλές διαφορετικές αρχιτεκτονικές CNN, τα βασικά τους γνωρίσματα είναι πανομοιότυπα. Ένα τυπικό CNN αποτελείται από μία στοιβάδα διαφορετικών επιπέδων όπως τα συνελκτικά επίπεδα (convolutional layers), τα συγκεντρωτικά επίπεδα (pooling layers) και τα πλήρως συνδεδεμένα επίπεδα (fully connected layers). Η απόδοση ενός CNN μοντέλου, που λειτουργεί υπό συγκεκριμένους πυρήνες και βάρη, υπολογίζεται με τη χρήση μιας συνάρτησης απώλειας μέσω της μεθόδου forward propagation σε ένα σύνολο δεδομένων εκπαίδευσης, ενώ οι παράμετροι που εκπαιδεύονται όπως οι πυρήνες και τα βάρη ενημερώνονται σύμφωνα με την τιμή απώλειας μέσω της μεθόδου Back Propagation και του αλγόριθμου βελτιστοποίησης Gradient Descent.



Εικόνα 3 η δομή των επιπέδων ενός CNN.

Η αρχιτεκτονική ενός τυπικού CNN με convolutional layers, pooling layers, fully connected layers (FC) και rectified linear units (ReLU) [26].

Το συνελκτικό επίπεδο (convolution layer) είναι η καρδιά της αρχιτεκτονικής ενός CNN, αφού είναι το επίπεδο στο οποίο διεξάγεται η εξαγωγή των χαρακτηριστικών των δεδομένων εισόδου και αποτελείται από ένα συνδυασμό γραμμικών και μη γραμμικών διεργασιών όπως η συνέλιξη και η συνάρτηση ενεργοποίησης.

Το συγκεντρωτικό επίπεδο (pooling layer) είναι το επίπεδο που διεξάγεται η δειγματοληψία κατά την οποία μειώνεται η διαστασιοποίηση των ανά επίπεδο χαρτογραφημένων χαρακτηριστικών (feature maps) με απώτερο σκοπό τη μείωση των εκπαιδευόμενων παραμέτρων που ακολουθούν. Αξίζει να σημειώσουμε ότι στο συγκεντρωτικό επίπεδο (pooling layer) δεν υπάρχουν παράμετροι προς εκπαίδευση και τον αντίστοιχο ρόλο τον διαδραματίζουν υπερ-παράμετροι όπως το μέγεθος του φίλτραρίσματος και το βήμα.

Τα τελικά χαρτογραφημένα χαρακτηριστικά μετατρέπονται σε ένα μονοδιάστατο πίνακα αριθμών και συνδέονται σε ένα ή περισσότερα πλήρως συνδεδεμένα επίπεδα (fully-connected layers), στα οποία η κάθε είσοδος είναι συνδεδεμένη με κάθε έξοδο μέσω ενός εκπαιδευμένου βάρους. Στη συνέχεια τα χαρακτηριστικά που θα εξαχθούν από τα συγκεντρωτικά επίπεδα (convolutional layers) και θα δειγματοληφθούν από τα συγκεντρωτικά επίπεδα (pooling layers), θα χαρτογραφηθούν από ένα υποσύνολο των πλήρως συνδεδεμένων επιπέδων (fully connected layers) στις τελικές εξόδους του δικτύου. Ένα παράδειγμα αυτής της διαδικασίας είναι οι πιθανότητες της κάθε κλάσης σε προβλήματα ταξινόμησης. Το τελευταίο πλήρως συνδεδεμένο επίπεδο (fully connected layer) συνήθως έχει τον ίδιο αριθμό κόμβων εξόδου με των αριθμό των κλάσεων, ενώ κάθε ένα από αυτά ακολουθείται συνήθως από μία μη γραμμική συνάρτηση όπως τα ReLU [27].

Η σχέση που υπολογίζει μαθηματικά την αξία ενός χαρακτηριστικού στη θέση (i,j) στο k-στο χαρτογραφημένο χαρακτηριστικό του l-στου επιπέδου, $z_{i,j,k}^l$ είναι η παρακάτω:

$$z_{i,j,k}^l = w_k^l \cdot x_{i,j}^l + b_k^l$$

Με τα w_k^l και b_k^l να συμβολίζουν το διανυσματικό βάρος και το σφάλμα του k-στου φίλτρου στο l-στο επίπεδο και με το $x_{i,j}^l$ την παρτίδα των δεδομένων εισόδου που τοποθετημένα στη θέση (i,j) του l-στου επιπέδου.

Η αξία ενεργοποίησης $a_{i,j,k}^l$ του συνελκτικού (convolutional) χαρακτηριστικού $z_{i,j,k}^l$ μπορεί να υπολογιστεί με την παρακάτω μαθηματική σχέση: $a_{i,j,k}^l = a(z_{i,j,k}^l)$. Με τις συνηθέστερες συναρτήσεις ενεργοποίησης να είναι οι σιγμοειδείς, tanh και ReLU. Η μείωση της διαστασιοποίησης που διεξάγεται στο συγκεντρωτικό επίπεδο (pooling layer) υπολογίζεται με την ενεργοποίηση της συγκεντρωτικής συνάρτησης pool για κάθε χαρτογραφημένο χαρακτηριστικό $a_{i,j,k}^l$ και έχει την παρακάτω μορφή:

$$y_{i,j,k}^l = pool(a_{m,n,k}^l), \forall (m,n) \in R_{ij}$$

Με το R_{ij} να συμβολίζει την γειτονικό χώρο γύρω από τη θέση (i,j). Τέλος αν υποθέσουμε ότι έχουμε N επιθυμητούς συσχετισμούς εισόδου-εξόδου $\{(x^{(n)}, y^{(n)})\}$, $n[1, \dots, N]$ με $x^{(n)}$ να συμβολίζεται το n-στο δεδομένο εισόδου, με $y^{(n)}$ να συμβολίζεται η αντίστοιχη τιμή στόχου, με $o^{(n)}$ να συμβολίζεται το αποτέλεσμα εξόδου και με το θ να ενεργοποιούνται όλες οι παράμετροι του CNN (βάρη, σφάλματα κλπ.), τότε η συνάρτηση απώλειας του CNN μπορεί να υπολογιστεί ακολούθως [27]:

$$L = \frac{1}{N} \sum_{n=1}^N l(\theta; y^{(n)}, o^{(n)})$$

3.4 Mask RCNN

Το Mask R-CNN είναι εννοιολογικά απλό: Το Faster R-CNN έχει δύο εξόδους για κάθε υποψήφιο αντικείμενο, μια επισήμανση και ένα πλαίσιο οριοθέτησης. Σε αυτό προστίθεται και έναν τρίτο κλαδί που εξάγει τη μάσκα αντικειμένου. Το Mask R-CNN είναι έτσι μια φυσική και διαισθητική ιδέα. Αλλά η πρόσθετη μάσκα εξόδου είναι ευδιάκριτη από την κλάση και τα πλαίσια οριοθέτησης εξόδου, που απαιτούν την εξαγωγή πολύ λεπτότερου χωρικού σχεδιαγράμματος ενός αντικειμένου. Στη συνέχεια, εισάγονται τα βασικά στοιχεία του Mask R-CNN, συμπεριλαμβανομένης της ευθυγράμμισης pixel-to-pixel, η οποία είναι το κύριο κομμάτι που λείπει από το Fast/Faster R-CNN.

Faster R-CNN: Αναθεωρώντας εν συντομία τον ταχύτερο ανιχνευτή R-CNN [28]. Το Faster R-CNN αποτελείται από δύο στάδια. Το πρώτο στάδιο, που ονομάζεται Region Proposal Network (RPN), προτείνει υποψήφια πλαίσια οριοθέτησης αντικειμένων. Το δεύτερο στάδιο, το οποίο είναι στην ουσία Fast R-CNN [29], εξάγει χαρακτηριστικά χρησιμοποιώντας RoIPool από κάθε υποψήφιο πλαίσιο και εκτελεί κατηγοριοποίηση και παλινδρόμηση για τα πλαίσια οριοθέτησης. Τα χαρακτηριστικά που χρησιμοποιούνται και από τα δύο στάδια μπορούν να μοιραστούν για ταχύτερα συμπεράσματα.

Μάσκα R-CNN: Το Mask R-CNN υιοθετεί την ίδια διαδικασία των δύο σταδίων, με ένα πανομοιότυπο πρώτο στάδιο (το οποίο είναι RPN). Στο δεύτερο στάδιο, παράλληλα με την πρόβλεψη της κλάσης και της μετατόπισης πλαισίων, το Mask R-CNN εξάγει επίσης μια δυαδική μάσκα για κάθε RoI. Αυτό έρχεται σε αντίθεση με τα πιο πρόσφατα συστήματα, όπου η κατηγοριοποίηση εξαρτάται από τις προβλέψεις της μάσκας (π.χ. [[30], [31], [32]]). Η προσέγγισή αυτή ακολουθεί την ιδέα του Fast R-CNN [29] που εφαρμόζει παλινδρόμηση και κατηγοριοποίηση των πλαισίων παράλληλα (η οποία αποδείχθηκε ότι απλοποιούσε σε μεγάλο βαθμό την σειρά πολλαπλών σταδίων του αρχικού R-CNN [[33]]).

Τυπικά, κατά τη διάρκεια της εκπαίδευσης, ορίζεται μια απώλεια πολλαπλών καθηκόντων σε κάθε δείγμα RoI ως $L = L_{cls} + L_{box} + L_{mask}$. Η απώλεια ταξινόμησης L_{cls} και η απώλεια οριοθέτησης L_{box} είναι ίδιες με εκείνες που ορίζονται στο [29]. Ο κλάδος ο οποίος είναι υπεύθυνος στο να εντοπίσει την μάσκα έχει μια έξοδο Km^2 - διαστάσεων για κάθε RoI, το οποίο κωδικοποιεί K δυαδικές μάσκες ανάλυσης $m \times m$, μία για κάθε K κλάση. Σε αυτό εφαρμόζουμε ένα sigmoid ανά εικονοστοιχείο, και ορίζουμε L_{mask} ως τη μέση δυαδική απώλεια διαγώνιος-εντροπίας (cross-entropy loss). Για ένα RoI που σχετίζεται με την αληθή κλάση k , το L_{mask} ορίζεται μόνο στη μάσκα στην k θέση (άλλες έξοδοι της μάσκας δεν συμβάλλουν στην απώλεια).

Ο ορισμός μας για το L_{mask} επιτρέπει στο δίκτυο να παράγει μάσκες για κάθε κατηγορία χωρίς ανταγωνισμό μεταξύ των κλάσεων. Βασιζόμενοι στον αποκλειστικό κλάδο κατηγοριοποίησης για να προβλέψουμε την κλάση που χρησιμοποιείται για να επιλεγεί η μάσκα εξόδου. Αυτό αποσυνδέει τη μάσκα από την πρόβλεψη της κλάσης. Αυτό είναι διαφορετικό από την κοινή πρακτική που εφαρμόζεται στα FCNs [34] στη σημασιολογική κατάτμηση (semantic segmentation), η οποία συνήθως χρησιμοποιεί ένα ανά-εικονοστοιχείο softmax και μια πολυωνυμική διαγώνια απώλεια εντροπίας. Σε αυτή την περίπτωση, οι μάσκες σε όλες τις κλάσεις ανταγωνίζονται. Στην περίπτωση του Mask R-CNN όμως, με ένα sigmoid ανά pixel και μια δυαδική απώλεια, δεν ανταγωνίζονται.

Αναπαράσταση μάσκας: Μια μάσκα κωδικοποιεί τη χωρική διάταξη ενός αντικειμένου εισόδου. Κατά συνέπεια και αντίθετα με τις ετικέτες κλάσης ή τα πλαίσια που αναπόφευκτα καταρρέουν στα μικρά διανύσματα εξόδου από τα πλήρως συνδεδεμένα (fully-connected) επίπεδα, η εξαγωγή της χωρικής δομής των μάσκων μπορεί να αντιμετωπιστεί φυσικά από την εικονοστοιχείο-σε- εικονοστοιχείο (pixel-to-pixel) αντιστοιχία που προέρχεται από τις συνελίξεις.

Συγκεκριμένα, προβλέπεται μια μάσκα $m \times m$ από κάθε RoI χρησιμοποιώντας ένα FCN [34]. Αυτό επιτρέπει σε κάθε επίπεδο στον κλάδο της μάσκας να διατηρεί τη ολόκληρη την $m \times m$ χωρική διάταξη αντικειμένου χωρίς να το καταστρέφει σε ένα διάνυσμα αναπαράστασης που στερείται χωρικών διαστάσεων. Σε αντίθεση με τις προηγούμενες μεθόδους που καταφεύγουν σε fc επίπεδα για την πρόβλεψη μάσκας [[30], [35], [31]], η πλήρης αναπαράσταση συνελίξης απαιτεί λιγότερες παραμέτρους, και είναι πιο ακριβής, όπως αποδεικνύεται από τα πειράματα.

Αυτή η εικονοστοιχείο-σε-εικονοστοιχείο (pixel-to-pixel) συμπεριφορά απαιτεί τα RoI χαρακτηριστικά, τα οποία είναι μικροί χάρτες χαρακτηριστικών γνωρισμάτων, να είναι καλά

ευθυγραμμισμένα για να διατηρήσουν πιστά τη ρητή χωρική αντιστοιχία ανά εικονοστοιχείο. Αυτό ήταν και ο λόγος που παρακίνησε να αναπτυχθεί το ακόλουθο επίπεδο RoIAlign που παίζει βασικό ρόλο στην πρόβλεψη μάσκας.

RoIAlign: Το RoIPool [29] είναι μια τυποποιημένη λειτουργία για την εξαγωγή ενός μικρού χάρτη χαρακτηριστικών γνωρισμάτων (π.χ., 7×7) από κάθε RoI. Το RoIPool ποσοτικοποιεί πρώτα έναν κυμαινόμενο αριθμό RoI στην διακριτή ευαισθησία του χάρτη χαρακτηριστικών γνωρισμάτων, αυτό το ποσοτικοποιημένο RoI έπειτα υποδιαιρείται σε χωρικές περιοχές οι οποίες ποσοτικοποιούνται, και τέλος συγκεντρώνονται οι τιμές των χαρακτηριστικών γνωρισμάτων που καλύπτονται από κάθε περιοχή (συνήθως από την ανώτατη συγκέντρωση). Η ποσοτικοποίηση πραγματοποιείται, π.χ., σε μια συνεχή συντεταγμένη x υπολογίζοντας $[x/16]$, όπου το 16 είναι ένας διασκελισμός χάρτη χαρακτηριστικών γνωρισμάτων και $[\cdot]$ είναι στρογγυλοποίηση. Ομοίως, η ποσοτικοποίηση πραγματοποιείται κατά τη διαίρεση σε περιοχές (π.χ. 7×7). Αυτές οι ποσοτικοποιήσεις εισάγουν τις αποκλίσεις μεταξύ του RoI και των εξαγόμενων χαρακτηριστικών γνωρισμάτων. Αν και αυτό μπορεί να μην επηρεάζει την κατηγοριοποίηση, η οποία είναι ισχυρή σε μικρές μεταφράσεις, αυτό έχει μεγάλη αρνητική επίδραση στην πρόβλεψη μασκών με ακρίβεια σε επίπεδο εικονοστοιχείου.

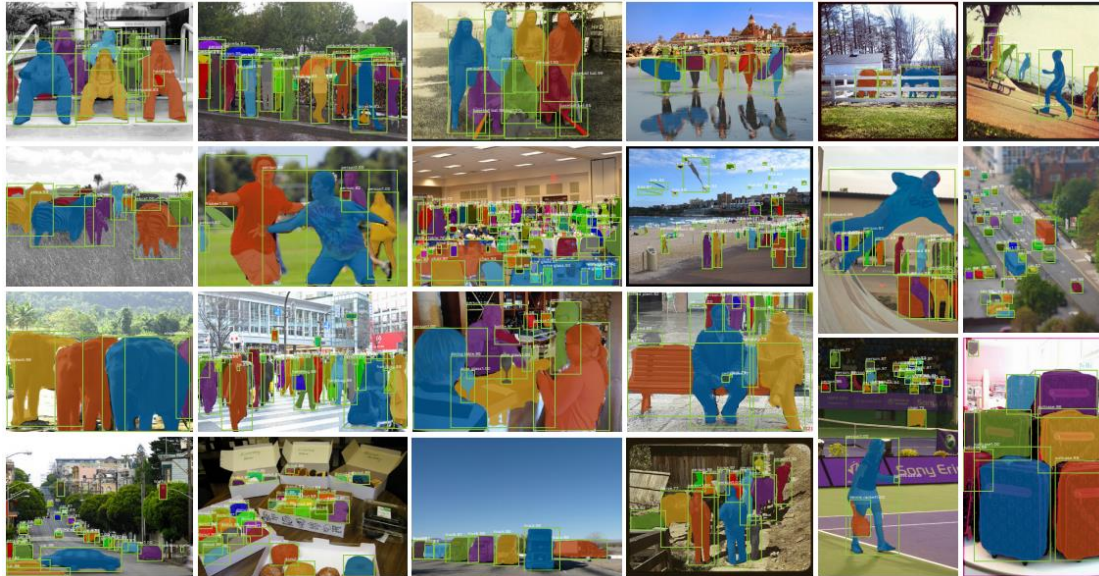
Για την αντιμετώπιση αυτού, προτείνουμε ένα επίπεδο RoIAlign που αφαιρεί τη σκληρή ποσοτικοποίηση του RoIPool, ευθυγραμμίζοντας σωστά τα χαρακτηριστικά που εξάγονται με την είσοδο. Η προτεινόμενη αλλαγή μας είναι απλή: αποφεύγουμε οποιαδήποτε ποσοτικοποίηση των RoI ορίων ή περιοχών (δηλαδή, χρησιμοποιούμε $x/16$ αντί του $[x/16]$). Χρησιμοποιούμε διγραμμική παρεμβολή [36] για να υπολογίσουμε τις ακριβείς τιμές των χαρακτηριστικών εισόδου σε τέσσερις τακτικά δειγματοληπτικές θέσεις σε κάθε RoI περιοχή και να συγκεντρώσουμε το αποτέλεσμα (χρησιμοποιώντας το μέγιστο ή μέσο όρο). Σημειώνουμε ότι τα αποτελέσματα δεν είναι ευαίσθητα στις ακριβείς θέσεις δειγματοληψίας ή στον αριθμό των σημείων που λαμβάνονται από το δείγμα, εφόσον δεν πραγματοποιείται ποσοτικοποίηση.

Το RoIAlign οδηγεί σε μεγάλες βελτιώσεις, το συγκρίνουμε επίσης και με τη λειτουργία RoIWarp που προτείνεται στο [31]. Αντίθετα από το RoIAlign, το RoIWarp αγνοεί το θέμα της ευθυγράμμισης και εφαρμόστηκε και στο [31] καθώς ποσοτικοποίησε το RoI ακριβώς όπως το RoIPool. Έτσι, ακόμη και αν το RoIWarp υιοθετεί επίσης διγραμμική αναδειγματοληψία παρακινούμενο από το [36], εκτελεί στο ίδιο επίπεδο με το RoIPool, όπως φαίνεται και από τα ανάλογα πειράματα, αποδεικνύοντας τον κρίσιμο ρόλο της ευθυγράμμισης.

Αρχιτεκτονική δικτύου: Για να αναδειχθεί η γενικότητα της προσέγγισής του Mask R-CNN, αρχικοποιείται με πολλαπλές αρχιτεκτονικές. Για λόγους σαφήνειας, διαφοροποιείται μεταξύ: (i) της backbone αρχιτεκτονικής της συνέλιξης που χρησιμοποιείται για την εξαγωγή χαρακτηριστικών σε μια ολόκληρη εικόνα και (ii) το δίκτυο για την αναγνώριση οριοθέτησης (κατηγοριοποίηση και παλινδρόμηση) και της πρόβλεψης μάσκας που εφαρμόζεται ξεχωριστά σε κάθε RoI.

Δηλώνοντας την αρχιτεκτονική του backbone χρησιμοποιώντας την ονοματολογία των network-depth -χαρακτηριστικών. Αξιολογώντας τα δίκτυα ResNet [37] και ResNeXt [38] βάθους 50 ή 101 επιπέδων. Η αρχική εφαρμογή του Faster R-CNN με ResNets [37] κατάφερε να εξάγει χαρακτηριστικά από το τελικό επίπεδο συνέλιξης του 4^{ου} σταδίου, το οποίο ονομάζεται C4. Αυτό το backbone με resNet-50, για παράδειγμα, συμβολίζεται από resNet-50-C4. Αυτή είναι μια κοινή επιλογή που χρησιμοποιείται στο [37], [31], [39], [40]]. Ένα ακόμη backbone που διερευνήθηκε στα πλαίσια του Mask R-CNN είναι το Feature Pyramid Network (FPN) [41]. Το FPN χρησιμοποιεί μια αρχιτεκτονική από επάνω προς τα κάτω με πλευρικές συνδέσεις για τη δημιουργία μιας πυραμίδας δυνατοτήτων εντός δικτύου από μια είσοδο μίας κλίμακας. Ο Faster R-CNN με μια FPN backbone εξάγει RoI χαρακτηριστικά από διαφορετικά

επίπεδα της πυραμίδας χαρακτηριστικών ανάλογα με την κλίμακα τους, κατά τα άλλα η υπόλοιπη προσέγγιση είναι παρόμοια με το πρωτότυπο ResNet. Χρησιμοποιώντας ένα ResNet-FPN backbone για την εξαγωγή χαρακτηριστικών με Mask R-CNN δίνει εξαιρετικά κέρδη τόσο στην ακρίβεια όσο και στην ταχύτητα.



Εικόνα 4. Μερικά αποτελέσματα του Mask R-CNN σε δοκιμαστικές εικόνες από το COCO, χρησιμοποιώντας ResNet-101-FPN με 35.7 mask AP Εικόνα από [42]].

	backbone	AP	AP50	AP75	APS	APM	APL
MNC [10]	ResNet-101-C4	24.6	44.3	24.8	4.7	25.9	43.6
FCIS [26] +OHEM	ResNet-101-C5-dilated	29.2	49.5	-	7.1	31.3	50.0
FCIS+++ [26] +OHEM	ResNet-101-C5-dilated	33.6	54.5	-	-	-	-
Mask R-CNN	ResNet-101-C4	33.1	54.9	34.8	12.1	35.6	51.1
Mask R-CNN	ResNet-101-FPN	35.7	58.0	37.8	15.5	38.1	52.4
Mask R-CNN	ResNeXt-101-FPN	37.1	60.0	39.4	16.9	39.9	53.5

Πίνακας 1. Διαχωρισμός παραδειγμάτων με Mask R-CNN στο COCO. MNC και FCIS είναι οι νικητές του COCO 2015 και 2016 διαγωνισμού αναλόγως. Όπως φαίνεται ο Mask R-CNN είναι πολύ πιο ακριβής από το πολύ σύνθετο FCIS+++ , το οποίο εμπειριέχει και πολλαπλές κλίμακες σε δεδομένα εκπαίδευσης/ελέγχου, οριζόντια αναστροφή στα δεδομένα ελέγχου και OHEM [43]]. Όλες οι δοκιμές είναι από ένα μοντέλο αποτελέσματα. Πίνακας από [42]].

3.5 Μετρικές αξιολόγησης

3.5.1 Mean Average Precision

Ο δείκτης ο οποίος χρησιμοποιείται προκειμένου να αξιολογήσουμε την απόδοση του μοντέλου ονομάζεται μέσος μέσης ακρίβειας (mAP). Η μέση ακρίβεια (AP) είναι μία δημοφιλής μετρική για την μέτρηση της ακρίβειας ανιχνευτών αντικειμένων όπως τα Faster R-CNN, SSD. Η μέση ακρίβεια υπολογίζει την μέση τιμή ακριβείας για την τιμή της ανάκλησης (Recall) από 0 έως 1.

3.5.2 Ακρίβεια και ανάκληση

Το πρόβλημα που προσεγγίζεται έχει δύο βασικές περιοχές, η πρώτη είναι να εντοπιστούν σωστά οι περιοχές στις εικόνες, το δεύτερο είναι να χαρακτηριστούν με μία κλάση οπότε ουσιαστικά πρόκειται για ένα πρόβλημα κατηγοριοποίησης. Σε προβλήματα κατηγοριοποίησης υπάρχουν και μετρικές όπως η ακρίβεια (accuracy), η ισορροπημένη ακρίβεια (balanced acc.), η ανάκληση (recall) που έχουν αξία και αφορούν την ακρίβεια στις προβλέψεις για την κλάση.

Ακρίβεια: μετράει το πόσο ακριβής είναι οι προβλέψεις. Ουσιαστικά το ποσοστό των προβλέψεων που είναι σωστές.

Ανάκληση: μετράει το πόσο καλά εντοπίζονται όλες οι θετικές προβλέψεις.

$TP = True\ positive$

$TN = True\ negative$

$FP = False\ positive$

$FN = False\ negative$

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

$$Recall = \frac{TP}{TP + FN}$$

$$Specificity = \frac{TN}{FP + TN}$$

$$Sensitivity = \frac{TP}{TP + FN}$$

$$Balanced\ acc. = \frac{Specificity + Sensitivity}{2}$$

3.5.3 IoU (Intersection over union)

Η μετρική αυτή μετράει την επικάλυψη δυο οριοθετημένων περιοχών μιας εικόνας. Αυτό χρησιμοποιείται για να υπολογιστεί το πόσο υπερκαλύπτονται μια περιοχή που έχει προβλεφθεί με την πραγματική περιοχή που έγινε πρόβλεψη. Στα περισσότερα σύνολα δεδομένων προκαθορίζεται ένα όριο με τιμή συνήθως 0,5, το οποίο χρησιμοποιείται προκειμένου να κατηγοριοποιηθεί μια πρόβλεψη σαν ορθά θετική ή λανθασμένα θετική.

3.6 Γλώσσα και βιβλιοθήκες προγραμματισμού

Οι γλώσσα προγραμματισμούς που επιλέχθηκε για την υλοποίηση του συγκεκριμένου αλγόριθμου είναι η Python 3.7.3. Η Python είναι μία πανίσχυρη γλώσσα που χρησιμοποιείται πολύ στον συγκεκριμένο τομέα της πληροφορικής διότι υπάρχει μια σειρά από εργαλεία που εξυπηρετούν τον τομέα αυτόν αλλά και την δυνατότητα που να παρέχει η γλώσσα αυτή να δημιουργούνται βιβλιοθήκες οι οποίες εσωτερικά τρέχουν C++ με αποτέλεσμα να έχουν πολύ μεγάλη απόδοση.

Η βασική βιβλιοθήκη που χρησιμοποιείται είναι το Mask RCNN και αυτή έχει υλοποιηθεί πάνω στο Tensorflow. Οπότε η υλοποίηση έχει στηριχθεί κυρίως στο πακέτο του Tensorflow (1.15), στο numpy για βασικές διεργασίες πάνω σε πίνακες, στο skimage για την υλοποίηση του SLIC και τέλος στο imgaug για περιστροφές και αναστροφές στις εικόνες προκειμένου να μεγαλώσουμε το όγκο των δεδομένων εκπαίδευσης.

Εκτός από τα παραπάνω πακέτα που χρησιμοποιούνται στον κορμό της υλοποίησης αυτής, επιπρόσθετα χρησιμοποιούνται και τα παρακάτω με πρόσθετες λειτουργίες όπως δημιουργία διαγραμμάτων ή διάβασμα σύνθετων αρχείων (πχ .h5 αρχεία).

Το σύνολο των python πακέτων που χρειάζονται προκειμένου να τρέχει η υλοποίηση αυτή, βρίσκονται στο requirements.txt αρχείο στον αρχικό φάκελο της εφαρμογής και αναφέρονται και στο [Παράρτημα IV](#).

4 Προτεινόμενη Υλοποίηση

4.1 Υποδομή και εργαλεία

4.1.1 Υποδομή

Το πρόγραμμα αυτό είναι ένα πρόγραμμα το οποίο εκτελεί μεγάλο πλήθος υπολογισμών, συνεπώς απαιτεί αρκετούς υπολογιστικούς πόρους για να τρέξει. Τα προγράμματα τα οποία εκπαιδεύουν μοντέλα τέτοιου είδους απαιτούν ειδικές κάρτες γραφικών για τρέξουν γρήγορα λόγω του μεγάλου αριθμού πυρήνων που διαθέτουν. Εκτός από υπολογιστική ισχύ χρειάζονται και αρκετά GB μνήμης γιατί ουσιαστικά πρόκειται επεξεργασία εικόνων και τα αρχεία φτάνουν σε μέγεθος μερικά MB. Επιπρόσθετα στην συγκεκριμένη υλοποίηση γίνονται και διαδικασίες όπως αναστροφές και περιστροφές των εικόνων, κάτι το οποίο γίνεται στην μνήμη με αποτέλεσμα οι ανάγκες για μνήμη να είναι σχετικά αυξημένες.

Για τους παραπάνω λόγους γίναν αρκετές δοκιμές σε διάφορα περιβάλλοντα προκειμένου να βρεθεί το βέλτιστο. Σε πρώτη φάση δοκιμάστηκε να τρέξει σε έναν υπολογιστή χρησιμοποιώντας την CPU (Intel i7-8850H, 2.6 GHz). Η δεύτερη δοκιμή έγινε σε μία GPU (Intel UHD Graphics 630, 192 πυρήνες, κοινή μνήμη με το λειτουργικό) και τέλος η δοκιμή με την μεγαλύτερη απόδοση ήταν να γίνει χρήση μίας GPU (NVIDIA TITAN Xp, 3840 πυρήνες, 12GB μνήμη). Το πρόγραμμα εκπαίδευσης που χρησιμοποιείται στις δοκιμές στην προσαρμογή των προβλέψεων καθώς και δοκιμές στις προσαρμοσμένες κλάσεις χρειάστηκε το παρακάτω χρόνο για κάθε περίπτωση υποδομής.

Υπολογιστική ισχύς	Χαρακτηριστικά	Πυρήνες	Χρόνος εκτέλεσης
CPU	Intel i7-8850H, 2.6 GHz	12	12h+
GPU	Intel UHD Graphics 630	192	7h
GPU	NVIDIA TITAN Xp	3840	2h 58min

Πίνακας 2 χρόνοι εκπαίδευσης νευρωνικού δικτύου.

4.1.2 Docker

Για την εύκολη εκτέλεση αλλά και για να είναι ανεξάρτητη η εφαρμογή από το λειτουργικό και τις ρυθμίσεις του υπολογιστή που θα τρέχει, έχει γίνει χρήση containers.

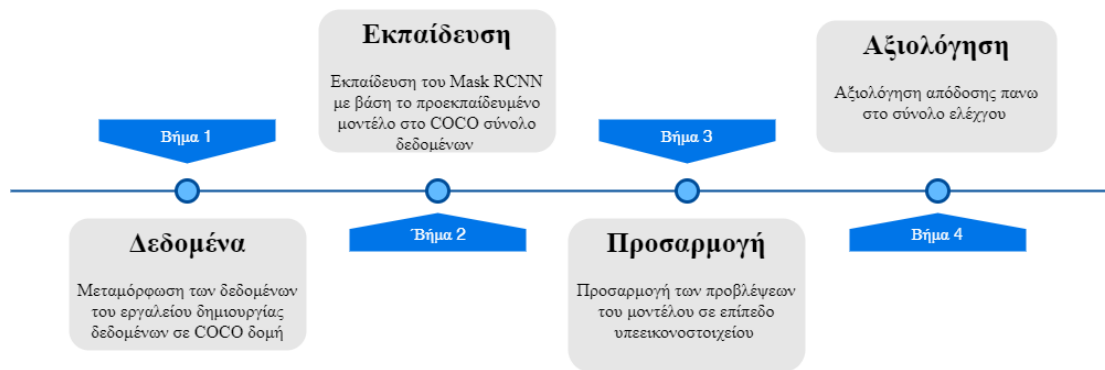
Τα Containers προσφέρουν πολλά πλεονεκτήματα όταν πρόκειται να χρησιμοποιηθούν σύνθετα περιβάλλοντα. Επιπλέον η χρήση της συγκεκριμένης τεχνολογίας διευκολύνει στην φορητότητα της εφαρμογής και δίνει την δυνατότητα να μπορεί να αναπτυχθεί σε διάφορα περιβάλλοντα με διαφορετικά λειτουργικά συστήματα και hardware.

Ουσιαστικά κάθε φορά που πρέπει να τρέξει η εφαρμογή δημιουργείται ένας container με συγκεκριμένο λειτουργικό σύστημα και ρυθμίσεις. Έτσι δίνεται η δυνατότητα στο σύστημα που φιλοξενεί το container (host machine) της εφαρμογής να δημιουργήσει ένα περιβάλλον το οποίο έχει μόνο τα εργαλεία τα οποία χρειάζεται προκειμένου να τρέξει βέλτιστα η εφαρμογή. Αυτό επιτυγχάνεται μέσω ενός προκαθορισμένου Dockerfile το οποίο περιέχει τις εντολές που χρειάζονται για να δημιουργηθεί το περιβάλλον που μόλις περιεγράφηκε.

4.2 Υλοποίηση

4.2.1 Σύνοψη υλοποίησης

Η υλοποίηση προκειμένου να δημιουργηθεί ένα μοντέλο το οποίο να χρησιμοποιεί τα δεδομένα που δημιουργούνται με το εργαλεία δημιουργίας δεδομένων και να κάνει προβλέψεις σε νέες εικόνες διακρίνεται στα παρακάτω βήματα.



Εικόνα 5. Η βασική δομή της υλοποίησης.

Προϋπόθεση για το παραπάνω είναι ότι έχει δημιουργηθεί ήδη ένα σύνολο δεδομένων από το εργαλείο δημιουργίας δεδομένων. Κατά την εκπαίδευση του μοντέλου υπήρχε διαθέσιμο ένα σύνολο δεδομένων οπότε ξεκινώντας με αυτό και όπως περιγράφεται στο παραπάνω διάγραμμα, σε πρώτη φάση θα πρέπει να μεταμορφωθούν τα δεδομένα από την μορφή εξόδου του εργαλείου, στην μορφή που χρησιμοποιεί και το COCO σύνολο δεδομένων [1]. Οι επισημάνσεις που έχουν γίνει στις εικόνες είναι σε επίπεδο υπερεικονοστοιχείου [44] οπότε οι επισημάνσεις είναι ουσιαστικά κάποια υπερεικονοστοιχεία που συνοδεύονται από την κλάση που χαρακτηρίστηκαν.

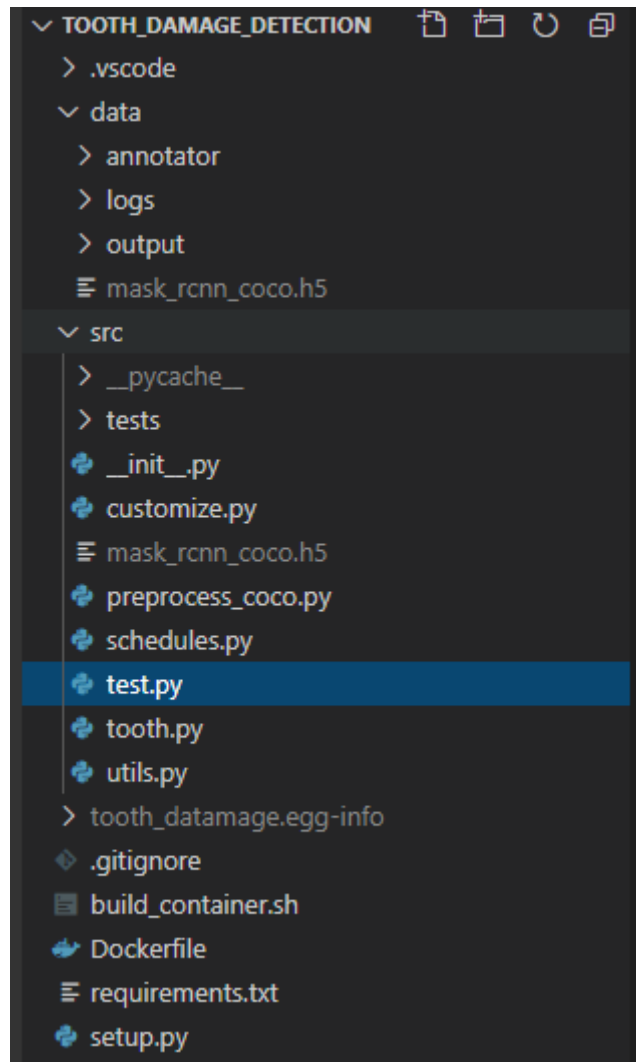
Στην συνέχεια, αφού τα δεδομένα έχουν έρθει στην μορφή που απαιτείται, ξεκινάει η εκπαίδευση του νευρωνικού χρησιμοποιώντας σαν βάση ένα μοντέλο που έχει ήδη εκπαιδευτεί με το σύνολο δεδομένων COCO.

Στο επόμενο βήμα το μοντέλο έχει εκπαιδευτεί σύμφωνα με το πρόγραμμα εκπαίδευσης και έχει προσαρμοστεί στο σύνολο δεδομένων που του έχει δοθεί και είναι σε θέση να κάνει προβλέψεις σε νέες εικόνες δοντιών. Σε αυτό το σημείο και στο επόμενο βήμα της υλοποίησης, γίνεται επέμβαση στις προβλέψεις που κάνει το μοντέλο σε επίπεδο εικονοστοιχείου και η εφαρμογή τις προσαρμόζει σε επίπεδο υπερεικονοστοιχείου. Με τον τρόπο αυτό η προβλέψεις έρχονται στην μορφή που δόθηκαν και κατά την εκπαίδευση.

Τέλος η εφαρμογή αξιολογεί τις προβλέψεις που έκανε το μοντέλο σε μία εικόνα συγκρίνοντάς τες με την επισήμανση που έχει γίνει από τους γιατρούς για την συγκεκριμένη εικόνα.

4.2.2 Δομή και αρχεία προγράμματος

Η εφαρμογή όπως αναφέρθηκε και νωρίτερα είναι φτιαγμένη σε python, οπότε ακολουθεί την παρακάτω δομή.



Εικόνα 6. Βασική δομή αρχείων του προγράμματος.

Με σειρά ο πρώτο φάκελος ο όνομα 'data' εμπεριέχει όλα τα δεδομένα που έχουν δημιουργηθεί με το εργαλείο δημιουργίας δεδομένων καθώς και το σύνολο των εικόνων. Τα δεδομένα αυτά έχουν χωριστεί σε τρεις ομάδες, που αντιστοιχούν σε έναν φάκελο η κάθε μία. Στον πρώτο φάκελο (training) υπάρχουν τα δεδομένα που χρησιμοποιούνται για να εκπαιδευτεί το μοντέλο. Ουσιαστικά είναι 72 εικόνες οι οποίες έχουν και ένα συνοδευτικό αρχείο .zip δεδομένα που δημιουργήθηκαν με το εργαλείο δημιουργίας δεδομένων. Αντιστοίχως στον δεύτερο υπάρχουν 12 εικόνες με τα συνοδευτικά αρχεία τα οποία χρησιμοποιούνται για την επικύρωση του μοντέλου κατά την διαδικασία της εκπαίδευσης. Τέλος υπάρχουν και 4 εικόνες στον τρίτο φάκελο οι οποίες είναι τελείως άγνωστες στο μοντέλο κατά την διαδικασία της εκπαίδευσης. Οι εικόνες αυτές χρησιμοποιούνται για να υπολογιστεί η πραγματική απόδοση του μοντέλου.

Ο επόμενος φάκελος της εφαρμογής(src) είναι ο φάκελος που περιέχει όλον τον κώδικα. Ο κώδικας είναι χωρισμένος σε τρία αρχεία έτσι ώστε να χωριστούν λογικά τα στάδια του προγράμματος.

Την επεξεργασία των δεδομένων την αναλαμβάνει το preprocess_coco.py. Η βασική ρουτίνα σε αυτό το αρχείο είναι το process_data το οποίο αναλαμβάνει να διαβάσει από έναν φάκελο όλα τα αρχεία με επισημάνσεις για πιθανές βλάβες και αφού τα αναλύσει δημιουργεί ένα νέο .json αρχείο το οποίο εμπεριέχει συγκεντρωμένη όλη την πληροφορία για τις περιοχές με βλάβη για την κάθε εικόνα. Επίσης δημιουργεί και αναθέτει ένα αναγνωριστικό σε όλες τις

εικόνας καθώς και στις κλάσεις. Η συγκεκριμένη δομή είναι ίδια με την δομή για τις επισήμανσεις που δίνεται και από το COCO σύνολο δεδομένων της Microsoft.

Το αρχείο `customize.py` εμπεριέχει όλη την λογική που έχει υλοποιηθεί προκειμένου να βελτιωθούν ακόμα περισσότερο τα αποτελέσματα του Mask RCNN. Οπότε στο αρχείο αυτό υπάρχουν οι ρουτίνες που δημιουργούν υπερεικονοστοιχεία με μία εικόνα και αναλύοντας το αποτέλεσμα του Mask RCNN δημιουργούν νέες προβλέψεις. Η βασική ρουτίνα έχει όνομα `transform_masks_to_superpixel` η οποία χρησιμοποιεί την `create_superpixels` προκειμένου να δημιουργήσει εκ νέου υπερεικονοστοιχεία στην εικόνα που αναλύεται και στην συνέχεια καλεί την `combine_masks_and_superpixels` η οποία με την σειρά της συνδυάζει το αποτέλεσμα του νευρωνικού με τα υπερεικονοστοιχεία και δημιουργεί νέα αποτελέσματα.

Τέλος το αρχείο `tooth.py` είναι το βασικό αρχείο του προγράμματος και είναι αυτό που αναλαμβάνει εντοχρηστρώσει τις διαδικασίες που εκτελούνται.

Το `requirements.txt` αρχείο εμπεριέχει όλα τα πακέτα που χρειάζεται η εφαρμογή προκειμένου να τρέξει και χρησιμοποιείται από το `Dockerfile` κατά την δημιουργία του περιβάλλοντος στο οποίο θα τρέξει η εφαρμογή.

Το `Dockerfile` χρησιμοποιείται προκειμένου να δημιουργηθεί το κατάλληλο περιβάλλον προκειμένου να τρέξει η εφαρμογή αξιοποιώντας σωστά το `hardware` πάνω στο οποίο τρέχει. Στο συγκεκριμένο `dockerfile` χρησιμοποιείται σαν βάση μια εικόνα η οποία έχει εγκατεστημένη `python` έκδοσης 3 και `Tensorflow-gpu` 1.13.1 έτσι ώστε να τρέχει χρησιμοποιώντας την κάρτα γραφικών. Ουσιαστικά πρόκειται για μία έκδοση Linux η οποία έχει προ-εγκατεστημένα τα παραπάνω εργαλεία. Στην συνέχεια η επόμενες εντολές που εκτελούνται αφορούν την μεταφορά του κώδικα και των δεδομένων στο `container` που θα δημιουργηθεί. Αφού γίνει και ενημέρωση των λιστών των πακέτων που είναι διαθέσιμα για το συγκεκριμένο λειτουργικό σύστημα (`apt-get update`) εγκαθίσταται `Git` γιατί θα χρειαστεί στην συνέχεια για να εγκατασταθούν σωστά όλα τα `python` πακέτα που χρειάζονται. Στο επόμενο βήμα εγκαθίστανται όλα τα `python` πακέτα χρησιμοποιώντας το `requirements.txt` αρχείο (`pip install --requirement requirements.txt`). Τέλος εγκαθίστανται και κάποιο οδηγό που χρειάζεται το `python` πακέτο `OpenCV` προκειμένου να λειτουργήσει σωστά ([Παράρτημα I](#)).

Τέλος έχει δημιουργηθεί και ένα βοηθητικό αρχείο κελύφους (`build_container.sh`) προκειμένου να αυτοματοποιηθεί η διαδικασία επαναδημιουργίας του `container` σε ένα περιβάλλον που τρέχει ήδη. Αυτό το αρχείο χρησιμοποιήθηκε για να εξοικονομηθεί χρόνο κατά την διάρκεια της ανάπτυξης του παραπάνω λογισμικού.

Ο κώδικας είναι διαθέσιμος και στην δημόσια βιβλιοθήκη που βρίσκεται στον σύνδεσμο https://github.com/chrisfilippis/tooth_damage_detection.

4.2.3 Αρχικοποίηση

Το πρόγραμμα αυτό δέχεται πολλές παραμέτρους όταν εκτελείται έτσι ο χρήστης έχει την δυνατότητα για παραμετροποίηση στην γραμμή εντολών. Ο χρήστης μπορεί να ορίσει που βρίσκονται οι φάκελοι με τα δεδομένα από το εργαλείο επισήμανσης δεδομένων, που να δημιουργηθούν οι φάκελοι με τα τελικά δεδομένα, το μονοπάτι του αρχικού μοντέλου για να ξεκινήσει η εκπαίδευση, την τοποθεσία για να σώζονται τα μοντέλα σε κάθε επανάληψη καθώς και το αν θέλει να ξεκινήσει την εκπαίδευση από την αρχή ή να συνεχίσει από το τελευταίο σωσμένο μοντέλο. Οι παράμετροι που δέχεται βρίσκονται στο [Παράρτημα II](#).

4.2.4 Προεπεξεργασία δεδομένων

Τα δεδομένα είναι χωρισμένα σε σύνολο εκπαίδευσης, σύνολο επικύρωσης και σύνολο ελέγχου, τα οποία είναι χωρισμένα και στους ανάλογους φακέλους (*training, validation, unknown*). Τα δεδομένα είναι χωρισμένα με τέτοιο τρόπο έτσι ώστε να υπάρχει ποικιλομορφία στα σύνολα που συμμετέχουν στην εκπαίδευση και να υπάρχουν δείγματα από όλες τις κλάσεις καθώς και δείγματα που απεικονίζουν δόντια τα οποία βρίσκονται σε όλες τις διαστάσεις της εικόνας.

Αφού το πρόγραμμα διαβάσει τις παραμέτρους που δόθηκαν από τον χρήστη, διαβάζει τα δεδομένα εισόδου και αφού αντιγράψει όλες τις εικόνες στους τελικούς φακέλους επεξεργασίας, δημιουργεί κατά ανάλογα .json αρχεία με την πληροφορία των επισημάνσεων. Σε αυτό το στάδιο γίνεται χρήση του module `preprocess_coco` που περιεγράφηκε παραπάνω.

Τα αρχικά δεδομένα που συλλέχθηκαν συνολικά ήταν ένα σύνολο από εικόνες και ένα συμπιεσμένο αρχείο για κάθε μια από αυτές το οποίο εμπεριείχε πληροφορίες για περιοχές με βλάβες.

Τα δεδομένα αυτά έπρεπε να επεξεργαστούν προκειμένου να έρθουν σε μία δομή που θα ήταν κατάλληλη για να φορτωθούν στο μοντέλο κατά την εκπαίδευση.

Για τον λόγο αυτό, στο πρόγραμμα που δημιουργήθηκε, το πρώτο στάδιο είναι η προσπέλαση όλων αυτών το συμπιεσμένων αρχείων προκειμένου να δημιουργηθεί ένα τελικό αρχείο σε μορφή JSON.

Η δομή του JSON αρχείου ακολουθεί την δομή του COCO (<http://cocodataset.org/>) συνόλου δεδομένων, ένα πολύ γνωστό σύνολο δεδομένων το οποίο χρησιμοποιείται σαν βάση, δεδομένου ότι το αρχικό μοντέλο που χρησιμοποιείται είναι ένα μοντέλο το οποίο έχει ήδη εκπαιδευτεί με το συγκεκριμένο σύνολο δεδομένων. Η δομή των δεδομένων είναι η παρακάτω, όπως περιγράφεται και στην επίσημη σελίδα του COCO (<http://cocodataset.org/#format-data>).

```

{
  "images": [
    {
      "id": 1,
      "width": 1024,
      "height": 768,
      "file_name": "antanastasio26.jpg",
      "date_captured": "2013-11-15 02:41:42"
    },
    {
      "id": 2,
      "width": 1024,
      "height": 768,
      "file_name": "antanastasio36.jpg",
      "date_captured": "2013-11-15 02:41:42"
    }
  ],
  "annotations": [
    {
      "id": 10000,
      "category_id": 2,
      "iscrowd": 0,
      "segmentation": [
        [
          297.0,
          161.0,
          298.0,
          160.0,
          298.0,
          159.0,
          299.0,
          158.0
        ]
      ]
    },
    {
      "id": 20000,
      "category_id": 1,
      "iscrowd": 0,
      "segmentation": [
        [
          208.0,
          128.0,
          209.0,
          128.0,
          210.0,
          127.0
        ]
      ]
    }
  ]
}

```

Εικόνα 7. Μερός δεδομένων σε μορφή COCO.

Πιο αναλυτικά υπάρχει η λίστα με όλες τις κλάσεις οι οποίες έχουν ένα χαρακτηριστικό και ένα όνομα, ακόμα υπάρχει μια λίστα με όλες τις εικόνες και πληροφορίες για αυτές όπως ύψος, πλάτος, αναγνωριστικό, όνομα και τέλος υπάρχει και μία λίστα με όλες τις περιοχές ενδιαφέροντος από τις εικόνες και με πληροφορίες για την κάθε περιοχή όπως σε ποια κλάση ανήκει, σε ποια εικόνα αναφέρεται το σύνολο των σημείων που ορίζουν την περιοχή αυτή κλπ.

Η συγκεκριμένη δομή είναι πολύ διαδομένη για περιπτώσεις τμηματοποίησης εικόνων, οπότε εξυπηρετεί και οποιονδήποτε στην συνέχεια θέλει να εκπαιδεύσει το συγκεκριμένο νευρωνικό με άλλα δεδομένα.

4.2.5 Παραμετροποίηση εκπαίδευσης

Στην συνέχεια φορτώνονται τα τρία σύνολα δεδομένων και τα αρχικοποιείται το μοντέλο.

Το μοντέλο χρησιμοποιεί σαν βάση ένα προεκπαιδευμένο μοντέλο το οποίο έχει ήδη εκπαιδευτεί πάνω στο σύνολο δεδομένων COCO. Στην συνέχεια ξεκινάει το πρόγραμμα [εκπαίδευσης](#) επιλέγοντας συγκεκριμένα επίπεδα του μοντέλου να αλλάξει.

Η διαμόρφωση που επιλέχτηκε για την εκπαίδευση καθώς και για την αξιολόγηση (απαιτείται να είναι η ίδια) είναι η ακόλουθη.

```
NAME = "coco"  
  
IMAGES_PER_GPU = 2  
  
NUM_CLASSES = 1 + 6  
  
STEPS_PER_EPOCH = 57  
  
RPN_ANCHOR_SCALES = (32, 64, 128, 256)  
  
IMAGE_MIN_DIM = 768  
  
IMAGE_MAX_DIM = 1024  
  
IMAGE_RESIZE_MODE = "none"  
  
VALIDATION_STEPS = 9
```

Αρχικά έχει δοθεί ένα όνομα στο συγκεκριμένο σύνολο ρυθμίσεων (NAME) και στην συνέχεια επιλέγεται το πλήθος των εικόνων (IMAGES_PER_CPU) που θα αναλύονται σε κάθε κάρτα γραφικών (GPU). Στην υποδομή που εκτελέστηκαν οι δοκιμές η κάρτα γραφικών ήταν μία με 12GB μνήμη, όμως επειδή οι εικόνες ήταν σχετικά υψηλής ανάλυσης ο αριθμός που επιλέχθηκε για την συγκεκριμένη ρύθμιση ήταν δύο. Αυτό προέκυψε ύστερα από αρκετές δοκιμές και δεν μπορούσε να μεγαλώσει άλλο γιατί κατά το βήμα εκπαίδευσης όλου του νευρωνικού το δίκτυο η μνήμη δεν ήταν αρκετή με αποτέλεσμα να διακόπτεται η εκπαίδευση.

Ο αριθμός κλάσεων στο συγκεκριμένο σύνολο δεδομένων είναι έξι οπότε στην ανάλογη ρύθμιση (NUM_CLASSES) μπήκε η τιμή επτά γιατί έπρεπε να συμπεριλάβουμε και το φόντο.

Στην συνέχεια τα βήματα ανά εποχή κατά την εκπαίδευση είναι 36. Αυτό προκύπτει γιατί το σύνολο δεδομένων για την εκπαίδευση είναι 72 δείγματα και δεδομένου ότι η κάρτα γραφικών επεξεργαζόταν δύο εικόνες κάθε φορά, τότε απαιτούνται 36 επαναλήψεις προκειμένου να αναλυθούν όλα.

Η επόμενη ρύθμιση που αλλάχτηκε είναι η RPN_ANCHOR_SCALES. Με αυτή την ρύθμιση, ρυθμίζουμε το RPN έτσι ώστε να δημιουργήσει τετραγωνικές άγκυρες με αυτές τις διαστάσεις σε εικονοστοιχεία.

Οι ρυθμίσεις που αφορούν το μέγεθος των φωτογραφιών εισόδου είναι οι IMAGE_MIN_DIM, IMAGE_MAX_DIM και IMAGE_RESIZE_MODE στις οποίες ορίζουμε κατά σειρά την ελάχιστη και μέγιστη δυνατή διάσταση καθώς και το αν υπάρχει ανάγκη για αλλαγή μεγέθους στις εικόνες εισόδου. Η επιλογή που έγινε εδώ είναι να μην γίνεται αλλαγή στο μέγεθος γιατί οι εικόνες ήταν μεγάλης ανάλυσης και θα μπορούσαν να έχουν περιοχές ενδιαφέροντος σε όλες τις περιοχές τους. Οπότε επιλέγεται να μην χαθεί καμία πληροφορία.

Τέλος η τελευταία ρύθμιση που άλλαξε είναι το πλήθος των βημάτων αξιολόγησης που θα τρέχει ο αλγόριθμος μετά το πέρας μίας εποχής (epoch).

4.2.6 Πρόγραμμα εκπαίδευσης

Το Mask RCNN είναι ένα νευρωνικό το οποίο δίνει την δυνατότητα κατά την εκπαίδευση να παγώσουμε κάποια επίπεδα έτσι ώστε να μην αλλοιωθεί η «γνώση» που έχει αποκτήσει από

προηγούμενη εκπαίδευση. Παρακάτω υπάρχουν οι προκαθορισμένες επιλογές που δίνονται από το δίκτυο για εκπαίδευση.

```
model.train(..., layers='heads', ...) # εκπαιδεύει τα τελικά στάδια (λιγότερη μνήμη)
```

```
model.train(..., layers='3+', ...) # εκπαιδεύει το resnet από το επίπεδο 3 και πάνω
```

```
model.train(..., layers='4+', ...) # εκπαιδεύει το resnet από το επίπεδο 4 και πάνω
```

```
model.train(..., layers='all', ...) # Train all layers (περισσότερη μνήμη)
```

Στην συγκεκριμένη υλοποίηση χρησιμοποιεί σαν βάση ένα προεκπαιδευμένο μοντέλο πάνω στο δημοφιλές σύνολο δεδομένων COCO και στην συνέχεια προσαρμόζονται κάποια επίπεδα του συγκεκριμένου έτσι ώστε να αποδίδει και στο νέο σύνολο δεδομένων.

Συνήθως στις περιπτώσεις όπως η παραπάνω, όταν υπάρχει ήδη ένα ισχυρό προεκπαιδευμένο μοντέλο, όταν γίνεται εκπαίδευση σε ένα νέο σύνολο δεδομένο επιλέγεται να αλλάξουν μόνο τα τελευταία επίπεδα του νευρωνικού, αυτά που είναι υπεύθυνα για το κατηγοριοποίηση των αποτελεσμάτων. Έτσι και στην παρούσα περίπτωση το πρόγραμμα εκπαίδευσης έχει ένα βήμα κατά το οποίο αλλάζουν μόνο τα τελευταία επίπεδα και στην συνέχεια ακολουθούν και άλλα βήματα στα οποία γίνονται και μερικές επαναλήψεις σε όλο το νευρωνικό ή και σε ένα μεγάλο μέρος του έτσι ώστε προσαρμοστεί και το υπόλοιπο δίκτυο.

Το πρόγραμμα το οποίο φάνηκε να αποδίδει περισσότερο από όλες τις δοκιμές που πραγματοποιήθηκαν είναι το παρακάτω ([Παράρτημα III](#)).

4.2.7 Αύξηση δεδομένων (Data Augmentation)

Εδώ αξίζει να σημειωθεί ότι λόγω περιορισμένου όγκου δεδομένων γίνεται χρήση μια τεχνικής που ονομάζεται augmentation προκειμένου να αυξήσουμε το πλήθος δειγμάτων που έχουμε για να κάνουμε εκπαίδευση. Αυτό επιτυγχάνεται επιλέγοντας τυχαίες αναστροφές ή περιστροφές των φωτογραφιών. Με τον τρόπο αυτό χρησιμοποιούνται και οι εικόνες όπως τραβήχτηκαν αλλά και οι ίδιες σε κάποιες παραλλαγές τους έχοντας προκύψει από τις παραπάνω διαδικασίες.

4.2.8 Συνδυασμός και προσαρμογή

Το πρόβλημα που καλείται να λύσει το μοντέλο είναι να εντοπίσει περιοχές σε δόντια που αντιστοιχούν σε συγκεκριμένες κλάσεις. Όμως πολλές φορές περιοχές δοντιών με βλάβη μιας κλάσης πολλές φορές μοιάζουν χρωματικά με γειτονικές περιοχές οι οποίες αντιστοιχούν σε βλάβες άλλης κλάσης. Αυτό είναι και το πραγματικό πρόβλημα που αντιμετωπίζουν πολλές φορές ο οδοντίατροι για τον λόγο αυτό επιστρατεύουμε μεθόδους από το πεδίο της μηχανικής μάθησης.

Το παραπάνω πρόβλημα είναι λογικό να δυσκολεύει και την εκπαίδευση του μοντέλου αλλά και να επηρεάζει αρνητικά και την απόδοσή του.

Με αφορμή το παραπάνω έγινε μια προσπάθεια για περαιτέρω ανάλυση των προβλέψεων του αλγόριθμου προκειμένου να προσεγγίσουμε όσο καλύτερα γίνεται την ακριβή διάγνωση μια οδοντικής βλάβης.

Αυτό πραγματοποιήθηκε με την χρήση και πάλι των υπερεικονοστοιχείων. Από την στιγμή που το σύνολο δεδομένων δημιουργήθηκε στηριζόμενο σε υπερεικονοστοιχεία θα ήταν λογικό να υπάρξει και μια πρόβλεψη σε αυτό το επίπεδο. Από την στιγμή που η πρόβλεψη από το Mask

RCNN γίνεται σε επίπεδο εικονοστοιχείου, έπρεπε η πληροφορία αυτή να μεταφερθεί σε επίπεδο υπερεικονοστοιχείου γιατί όπως αναφέρθηκε και νωρίτερα τα υπερεικονοστοιχεία αποτελούν ένα σύνολο εικονοστοιχείων που σημασιολογικά είναι πολύ κοντά. Οπότε η πληροφορία έπρεπε να κάνει ένα ταξίδι από το υπερεικονοστοιχείο, στο εικονοστοιχείο, στην συνέχεια στην ανάλυση και μετά αντίστροφα από το εικονοστοιχείο στο υπερεικονοστοιχείο.

Οπότε προκειμένου να ολοκληρωθεί αυτό ο κύκλος έπρεπε να μετατραπούν έξυπνα τα εικονοστοιχεία που προκύπταν σαν έξοδος από το μοντέλο σε υπερεικονοστοιχεία.

Για το συγκεκριμένο πρόβλημα έπρεπε να γενικευθεί η πληροφορία που υπήρχε σε επίπεδο εικονοστοιχείου σε επίπεδο υπερεικονοστοιχείου.

Για να γίνει αυτό έπρεπε σε πρώτη φάση να υπολογιστούν εκ νέου τα υπερεικονοστοιχεία που αντιστοιχούσαν στην εικόνα εισόδου και την συνέχεια έπρεπε να βρεθούν ποια εικονοστοιχεία που αντιστοιχούσαν σε κάθε υπερεικονοστοιχείο, έχουν συμπεριληφθεί στην πρόβλεψη του μοντέλου. Αν τελικά υπήρχαν εικονοστοιχεία στην πρόβλεψη τότε έπρεπε να βρεθεί η επικρατέστερη κλάση. Αυτό έγινε υπολογίζοντας το πλήθος των εικονοστοιχείων που αντιστοιχούσαν σε κάθε κλάση. Τέλος η κλάση με τα περισσότερα ήταν αυτή που επικρατούσε, οπότε κάνοντας την παραπάνω διαδικασία για όλα τα υπερεικονοστοιχεία προέκυπτε ένα νέο σύνολο προβλέψεων το οποίο πλέον αποτελείται από υπερεικονοστοιχεία. Οπότε συνοπτικά τα βήματα ήταν τα εξής:

- 1) Υπολογισμός υπερεικονοστοιχείων
- 2) Για κάθε υπερεικονοστοιχείο
 - i. Υπολογισμός των εικονοστοιχείων που συμπεριλαμβάνονται στις προβλέψεις του μοντέλου
 - ii. Υπολογισμός της επικρατέστερης κλάσης
 - iii. Δημιουργία νέας πρόβλεψης που αποτελείται από όλα τα σημεία του υπερεικονοστοιχείου και έχει κλάση την επικρατέστερη κλάση που υπολογίστηκε στο προηγούμενο βήμα.

Όσο αφορά το τεχνικό μέρος την συγκεκριμένη διεργασία την αναλαμβάνει εξολοκλήρου το python module `customize.py`. Η ρουτίνα που οργανώνει τις υπόλοιπες ρουτίνες είναι η `transform_masks_to_superpixel` η οποία πρώτα καλεί την `create_superpixels` η οποία δημιουργεί τα υπερεικονοστοιχεία και στην συνέχεια καλεί την `combine_masks_and_superpixels` η οποία δέχεται τις προβλέψεις του μοντέλου και τα υπερεικονοστοιχεία που δημιουργήθηκαν νωρίτερα. Εκεί για κάθε υπερεικονοστοιχείο υπολογίζεται το πλήθος των εικονοστοιχείων της κάθε κλάσης και αν υπάρχει έστω και μία κλάση με πλήθος μεγαλύτερο του μηδενός, τότε και μόνο τότε υπολογίζεται η επικρατέστερη κλάση και δημιουργείται μια νέα πρόβλεψη με το σχήμα του υπερεικονοστοιχείου και κλάση αυτή που μόλις υπολογίστηκε.

Αφού ολοκληρωθεί η παραπάνω διαδικασία για κάθε υπερεικονοστοιχείο τότε δημιουργείται ένα νέο σύνολο προβλέψεων το οποίο επιστρέφεται στο βασικό python module(`tooth.py`).

4.2.9 Αξιολόγηση

Τέλος το πρόγραμμα αφού ολοκληρώσει την εκπαίδευση προσπαθεί να αξιολογήσει την απόδοση του μοντέλου στα δεδομένα που δεν έχει χρησιμοποιήσει για να εκπαιδευτεί.

Την λειτουργία αυτή την αναλαμβάνει η `measure_accuracy` ο οποία κάνει προβλέψει πάνω στο άγνωστο σύνολο δεδομένων και υπολογίζει την απόδοση με την χρήστη μια μετρικής που ονομάζεται mAP (mean average precision).

Προσαρμογή

Αφού ο αλγόριθμος εκπαιδευτεί και λίγο πριν αξιολογήσει ένα αποτέλεσμα γίνεται μια προσαρμογή στην πρόβλεψη του μοντέλου. Αφού δώσουμε την εικόνα στο μοντέλο και επισημάνει κάποιες περιοχές με συγκεκριμένες κλάσεις και πλαίσια οριοθέτησης τότε χρησιμοποιούνται αυτά τα δεδομένα προκειμένου να γίνει μια καινούρια πρόβλεψη στηριζόμενη στο πρόβλεψη του νευρωνικού. Εκεί επεμβαίνει η ρουτίνα `transform_masks_to_superpixel` του `customize.py` η οποία χρησιμοποιώντας την `create_superpixels` δημιουργεί εκ νέου υπερεικονοστοιχεία κάνοντας χρήση της βιβλιοθήκης `skimage` και αποφασίζει πλέον ποια υπερεικονοστοιχεία πρέπει να επιστραφούν σαν αποτέλεσμα. Η λογική πίσω από την απόφαση αυτή περιγράφεται παρακάτω στην [ανάλογη ενότητα](#).

5 Πειραματική Αξιολόγηση και Αποτελέσματα

5.1 Δεδομένα

Τα δεδομένα για να εκπαιδευτεί το μοντέλο έπρεπε να ήταν επιστημονικά ακριβή και έπρεπε να έχουν συγκεκριμένη δομή προκειμένου να είναι εύκολα στην επεξεργασία. Συνεπώς έπρεπε να φωτογραφηθούν δόντια σε σχετική υψηλή ανάλυση προκειμένου να υπάρχει αρκετή πληροφορία για ανάλυση και στην συνέχεια έπρεπε να επέμβουν οι οδοντίατροι προκειμένου να αξιολογήσουν τις εικόνες και να επισημάνουν τις περιοχές με βλάβες πάνω σε αυτές.

Αυτό από μόνο του ήταν ένα δύσκολο εγχείρημα διότι έπρεπε να βρεθεί/δημιουργηθεί κατάλληλος εξοπλισμός αλλά και να αφιερωθεί χρόνος από γιατρούς προκειμένου να αξιολογήσουν τις εικόνες.

Τα δόντια που επιλέχθηκαν είχαν περιοχές οι οποίες είτε δεν είχαν βλάβες, είτε είχαν βλάβες οι οποίες μπορούσαν να κατηγοριοποιηθούν από το 1 έως το 6 του ICDAS II [45] συστήματος κατηγοριοποίησης. Στην κλίμακα αυτή προστέθηκε και μια επιπλέον κλάση, η κλάση 0 η οποία ουσιαστικά αντιπροσωπεύει μια υγιή περιοχή. Αυτή η κλάση θα μπορούσα να χαρακτηριστεί και ως φόντο (background). Η κλάση 0 συμπεριλαμβάνει όλες τις υγιείς περιοχές του δοντιού καθώς και τις περιοχές τις εικόνα οι οποίες δεν είναι οδοντική. Οπότε όταν μια περιοχή χαρακτηρίζεται με βλάβη κλάσης 1 σημαίνει ότι έχει μια πολύ μικρή βλάβη ενώ από την άλλη όσο μεγαλύτερη είναι η τιμή της κλίμακας σε μια περιοχή τόσο μεγαλύτερη είναι η βλάβη.

Δόντια με περιοχές με υποπλαστικές καθώς και με μεταλλικές περιοχές δεν συμπεριλαμβάνονται στο συγκεκριμένο σύνολο δεδομένων. Οι εικόνες που επιλέχθηκαν για την εκπαίδευση ήταν στο πλήθος 114 ενώ οι εικόνες επικύρωσης ήταν 18.

5.1.1 Διαδικασία λήψης εικόνας

Προκειμένου να δημιουργηθούν δεδομένα τα οποία θα χρησιμοποιηθούν για να εκπαιδευτεί το νευρωνικό δίκτυο, αρχικά έπρεπε να γίνουν κάποιες λήψεις πάνω σε δόντια ασθενών. Στην συνέχεια οι φωτογραφίες δόθηκαν σαν είσοδο στο εργαλείο χαρακτηρισμού εικόνων προκειμένου να δημιουργηθούν δεδομένα τα οποία χρησιμοποιήθηκαν για να ξεκινήσει η εκπαίδευση.

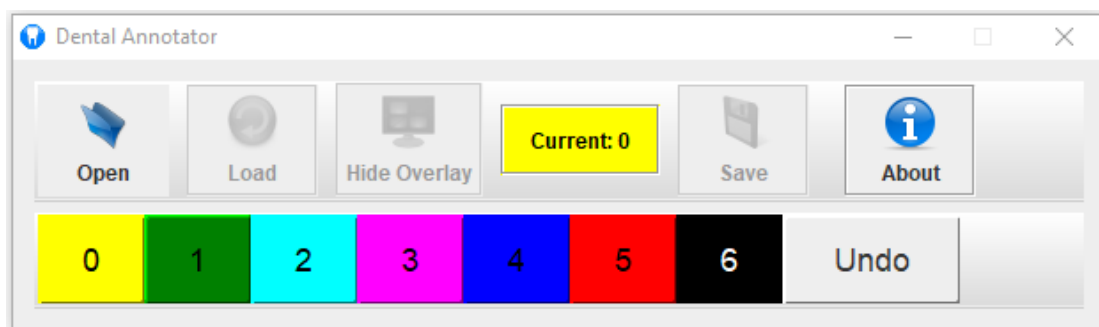
Τα δόντια που μελετήθηκαν φωτογραφήθηκαν με μία ενδοστοματική κάμερα. Αυτά πριν φωτογραφηθούν ακολούθησαν μια κοινή διαδικασία προετοιμασίας και καθαρισμού για πιο ευκρινείς φωτογραφίες.

Η διαδικασία καθαρισμού πραγματοποιήθηκε με μια τριχωτή βούρτσα με μικρή ταχύτητα. Στην συνέχεια τα δόντια ξεπλύθηκαν με νερό για δέκα δευτερόλεπτα και στην συνέχεια στέγνωσαν με αέρα για 5 επιπλέον δευτερόλεπτα. Τέλος αφού πλέον το επιλεγμένο δόντι ήταν πλέον καθαρό, φωτογραφήθηκε με μια ενδοστοματική κάμερα Carestream 1200 (Carestream Dental LLC, USA) με τον ίδιο προσανατολισμό σχετικά με το δόντι καθώς και με την ίδια απόσταση. Η διαστάσεις του αισθητήρα σε εικονοστοιχεία είναι 2592 (ύψος) x 1944 (πλάτος). Η ανάλυση του αισθητήρα κατά την διάρκεια της βιντεοσκόπησης σε αριθμό εικονοστοιχείων ήταν 640 (ύψος) x 480 (πλάτος) ενώ κατά την διάρκεια της φωτογράφισης η ανάλυση αυξήθηκε σε 1024 (ύψος) x 768 (πλάτος). Η εστιακή απόσταση ήταν από 3 χιλιοστά έως 25 και για τον φωτισμό υπήρχαν 6 LED λυχνίες. Η έξοδος της κάμερας ήταν τρεις διαφορετικές μορφές TV-NTSC, TV-PAL και S-Video. Η σύνδεση με τον υπολογιστή έγινε με την χρήση μια θύρας USB 2.0.

5.1.2 Εργαλείο χαρακτηρισμού εικόνων

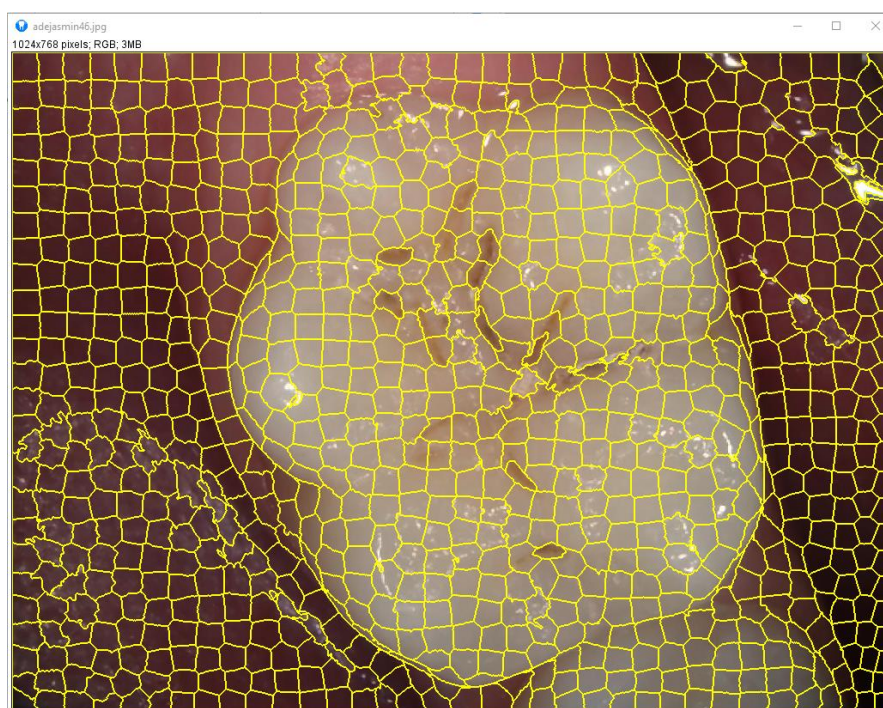
Αφού είχε διασφαλιστεί η ποιότητα των φωτογραφιών των δοντιών στην συνέχεια έπρεπε να επέμβει ο οδοντίατρος προκειμένου να χαρακτηρίσει τις περιοχές πάνω στα δόντια.

Για την παραπάνω ανάγκη δημιουργήθηκε ένα εργαλείο χαρακτηρισμού οδοντικών εικόνων. Το εργαλείο αυτό είναι γραμμένο σε JAVA και είναι σχετικά απλό στην χρήση. Ο χρήστης χρησιμοποιώντας το ανάλογο κουμπί (Open) φορτώνει την εικόνα, η οποία στην περίπτωση μας έχει τραβηχτεί νωρίτερα από την κάμερα που περιγράφηκε νωρίτερα, που επιθυμεί.



Εικόνα 8. Γραφική διεπαφή του εργαλείου δημιουργίας δεδομένων.

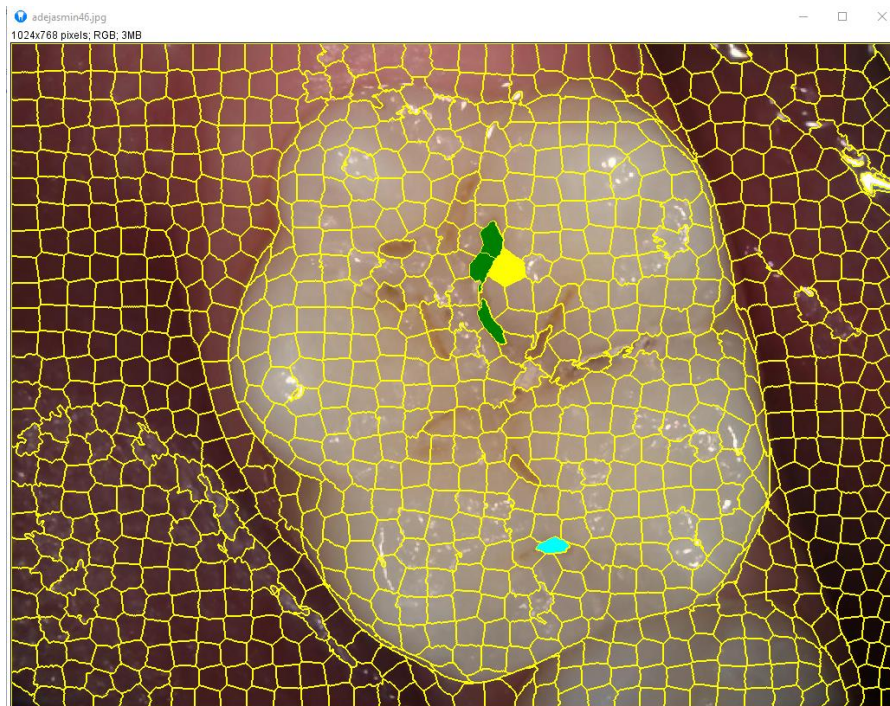
Στην συνέχεια το εργαλείο την επεξεργάζεται και την χωρίζει σε ομάδες γειτονικών εικονοστοιχείων χρησιμοποιώντας τον αλγόριθμο SLIC.



Εικόνα 9. Εικόνα έτοιμη για επισήμανση μέσω του εργαλείου δημιουργίας δεδομένων.

Αυτό βοηθάει γιατί η εικόνα χωρίζεται σε μικρότερα κομμάτια τα οποία έχουν παρόμοια χαρακτηριστικά. Έχοντας πλέον χωρισμένη την εικόνα σε κομμάτια μικρά αλλά όχι τόσο μικρό όσο ένα εικονοστοιχείο, ο χρήστης μπορεί να επιλέξει από την κλίμακα την τιμή που επιθυμεί

και να μαρκάρει την περιοχή που θέλει. Με τον τρόπο αυτό ο χρήστης του εργαλείου μπορεί να επιλέξει και να χαρακτηρίζει μέρη της εικόνας εύκολα και σχετικά γρήγορα.



Εικόνα 10. Εικόνα με μερικές επισημάνσεις.

Αφού ο χρήστης τελειώσει με τον εντοπισμό όλων των βλαβών που κρίνει ο ίδιος τότε μπορεί να επιλέξει το ανάλογο κουμπί (Save) προκειμένου να σώσει το μαρκάρισμα.

Το εργαλείο από την πλευρά του δημιουργεί ένα συμπιεσμένο αρχείο .zip το οποίο περιέχει αρχεία με όλες τις περιοχές που ομαδοποίησε χρησιμοποιώντας τον αλγόριθμο SLIC καθώς και τις επιλογές του χρήστη. Τα αρχεία αυτά είναι αρχεία .roi τα οποία περιέχουν το σχήμα της κάθε περιοχής σε μία σειρά από σημεία.

Name	Date modified	Type	Size
2.roi	10/12/2018 12:43 πμ	ROI File	1 KB
2-1.roi	10/12/2018 12:43 πμ	ROI File	2 KB
2-2.roi	10/12/2018 12:43 πμ	ROI File	1 KB
2-3.roi	10/12/2018 12:43 πμ	ROI File	1 KB
2-4.roi	10/12/2018 12:43 πμ	ROI File	1 KB
2-5.roi	10/12/2018 12:43 πμ	ROI File	1 KB
2-6.roi	10/12/2018 12:43 πμ	ROI File	1 KB
2-7.roi	10/12/2018 12:43 πμ	ROI File	1 KB
2-8.roi	10/12/2018 12:43 πμ	ROI File	1 KB
2-9.roi	10/12/2018 12:43 πμ	ROI File	1 KB
2-11.roi	10/12/2018 12:43 πμ	ROI File	1 KB
3.roi	10/12/2018 12:43 πμ	ROI File	2 KB
3-1.roi	10/12/2018 12:43 πμ	ROI File	1 KB

Εικόνα 11. Υπόδειγμα από τα αρχεία που περιέχονται σε ένα αρχείο επισήμανσης.

Με το παραπάνω εργαλείο οι οδοντίατροι μπόρεσαν να μαρκάρουν εικόνες υψηλής ανάλυσης, οι οποίες είχαν τραβηχτεί νωρίτερα από διάφορους ασθενείς και τα δεδομένα αυτά χρησιμοποιήθηκαν έτσι ώστε να εκπαιδευτεί το μοντέλο που περιγράφεται στην διατριβή αυτή.

5.2 Μετρικές απόδοσης Mask RCNN

Το συγκεκριμένο Νευρωνικό δίκτυο, σύμφωνα και τα χαρακτηριστικά που έχει στην έξοδο του, έχουν δημιουργηθεί κάποιες μετρικές οι οποίες αξιολογούν την απόδοση των επιμέρους χαρακτηριστικών και στο τέλος υπολογίζουν και την συνολική απόδοση προσθέτοντας τις παραπάνω επιμέρους μετρικές.

Οι επιμέρους μετρικές είναι πέντε στο πλήθος και αντιπροσωπεύουν το πόσο καλά το μοντέλο εντοπίζει χαρακτηριστικά σε μία εικόνα εισόδου. Οι δείκτες αυτοί υπολογίζονται και στο σύνολο δεδομένων εκπαίδευσης αλλά και στο σύνολο επικύρωσης και είναι οι παρακάτω.

`grn_class_loss` : αντιπροσωπεύει πόσο καλά το RPN ξεχωρίζει το φόντο με τα αντικείμενα.

`grn_bbox_loss` : αντιπροσωπεύει πόσο καλά το RPN εντοπίζει τα αντικείμενα

`mrcnn_bbox_loss` : πόσο καλά το Mask RCNN εντοπίζει τα αντικείμενα

`mrcnn_class_loss` : πόσο καλά το Mask RCNN αναγνωρίζει την κλάση του κάθε αντικειμένου

`mrcnn_mask_loss` : πόσο καλά το Mask RCNN ξεχωρίζει τα αντικείμενα

Επιπρόσθετα υπάρχουν και άλλοι δύο δείκτες οι οποίοι αφορούν συνολικά την απόδοση στα δεδομένα εκπαίδευσης (`loss`) και ο ανάλογος δείκτης για την απόδοση στα δεδομένα επικύρωσης.

5.3 Αποτελέσματα

Σε αυτή την ενότητα θα παρουσιαστούν αποτελέσματα από δοκιμές που πραγματοποιήθηκαν. Οι δοκιμές γίναν με σκοπό να βρεθούν διάφοροι τρόποι προκειμένου ο αλγόριθμος να κάνει προβλέψεις όσο το δυνατόν πιο κοντά στις επισημάνσεις που δόθηκαν από τους οδοντίατρος. Βαρύτητα δίδεται και τόσο στο αν μια πρόβλεψη καταφέρει να εντοπίσει την περιοχή στην εικόνα με την βλάβη αλλά και στο αν η πρόβλεψη είναι ακριβής όσο αφορά την κλάση.

Η ακρίβεια στην κλάση είναι σημαντική όμως δεδομένου ότι οι κλάσεις δεν αντιπροσωπεύουν διαφορετικά αντικείμενα ή καταστάσεις, αλλά το μέγεθος μια βλάβης σε ένα δόντι, οι μικρές αποκλίσεις σε γειτονικές κλάσεις είναι κάτι που αξιολογείται θετικά.

Αρχικά έγιναν κάποιες δοκιμές προκειμένου να βρεθεί το μοντέλο με την καλύτερη απόδοση. Σε αυτήν την σειρά δοκιμών έτρεξαν διαφορά προγράμματα εκπαίδευσης με διαφορετικές τιμές στις υπερ-παραμέτρους του νευρωνικού προκειμένου να βρεθεί το μοντέλο του οποίου οι προβλέψεις προσεγγίζουν καλύτερα τις επισημάνσεις των γιατρών.

Σε δεύτερο χρόνο επιλέγεται το μοντέλο με την μεγαλύτερη απόδοση και χρησιμοποιώντας τις προβλέψεις, οι οποίες γίνονται σε επίπεδο εικονοστοιχείων, γενικεύονται στα υπερεικονοστοιχεία που χρησιμοποιήθηκαν για να γίνει ο χαρακτηρισμός των εικόνων. Με τον τρόπο αυτό οι προβλέψεις γίνονται στο ίδιο επίπεδο με τους δοθέντες χαρακτηρισμούς.

Δηλαδή οι χαρακτηρισμοί δίνονται σε επίπεδο υπερεικονοστοιχείων, στην συνέχεια το Mask RCNN προσπαθεί να προσεγγίσει τις βλάβες σε μια εικόνα ενός δοντιού, η πρόβλεψή γίνεται σε επίπεδο εικονοστοιχείου, και τέλος εφαρμόζεται ένα αλγόριθμος ο οποίος προσαρμόζει τις

προβλέψεις και τις γενικοποιεί πίσω σε υπερεικονοστοιχεία όπως και οι χαρακτηρισμοί στην αρχή.

Τέλος για το σύνολο των έξι κλάσεων που χρησιμοποιούνται παρατηρούνται παρόμοια χαρακτηριστικά και βλάβες, ύστερα και από προτροπή των γιατρών έχει αξία να ενώσουμε τα αποτελέσματα κάποιων κλάσεων προκειμένου να δημιουργήσουμε τρεις κλάσεις πλέον με μεγαλύτερη ακρίβεια και εξίσου σημαντική πληροφορία ιατρικά.

5.3.1 Δοκιμές στο πρόγραμμα εκπαίδευσης

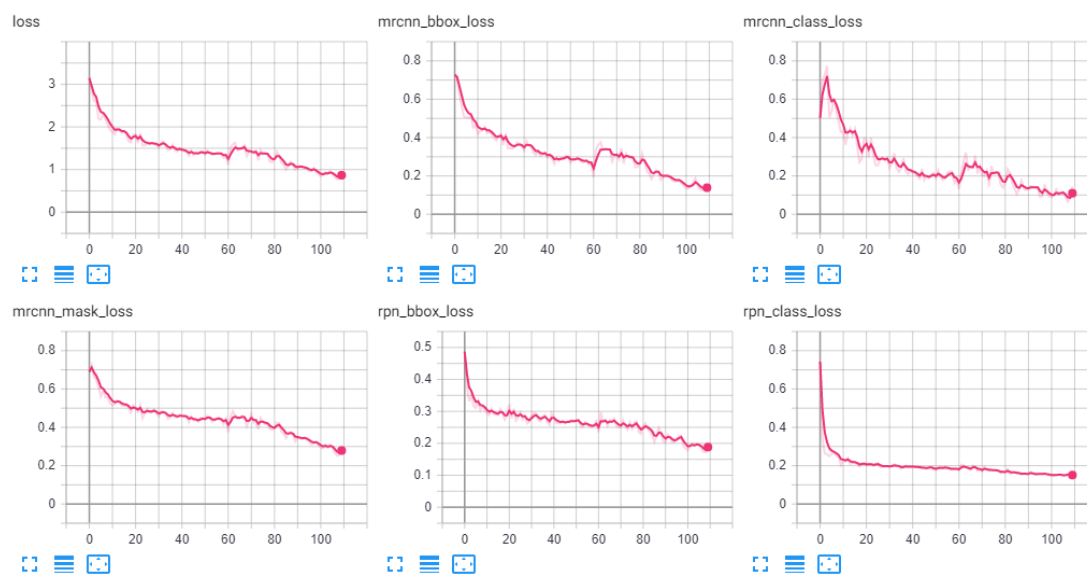
Η πρώτη σειρά από δοκιμές εστιάζει στα προγράμματα εκπαίδευσης και το πιο τελικά είναι πιο αποδοτικό. Όπως αναφέρθηκε και νωρίτερα κατά την εκπαίδευση δίνεται η δυνατότητα να επιλέγονται ποια του δικτύου για να παγώνουν κατά την εκπαίδευση. Με τον τρόπο αυτό αλλάζουν τα βάρη μόνο στα υπόλοιπα επίπεδα. Αυτό είναι χρήσιμο γιατί όταν χρησιμοποιείται προεκπαιδευμένο μοντέλο το οποίο έχει αποκτήσει γνώση από ένα πολύ μεγαλύτερο σύνολο δεδομένων, είναι ωφέλιμο στις περιπτώσεις που δεν υπάρχουν πολλά δεδομένα, όπως και στην παραπάνω περίπτωση.

5.3.1.1 Δοκιμή 1

Το πρώτο πρόγραμμα εκπαίδευσης που δοκιμάστηκε ήταν να ενημερωθούν μόνο τα τελευταία επίπεδα για 60 εποχές, στην συνέχεια να ενημερωθούν όλα τα επίπεδα το νευρωνικού για 50 εποχές. Ο ρυθμός μάθησης και στα δύο στάδια είχε τιμή 0.001. Επιπρόσθετα υπήρχε και augmentation στα δεδομένα κατά την εκπαίδευση με τυχαίες περιστροφές και αναστροφές 180 μοιρών.

Πρόγραμμα εκπαίδευσης:

- i. 60 εποχές ενημερώνοντας μόνο τα βάρη από τελευταία επίπεδα (heads)
- ii. 50 εποχές ενημερώνοντας όλα τα βάρη (all)



Εικόνα 12. Μετρικές απόδοσης στα δεδομένα εκπαίδευσης για την δοκιμή 1.

	Epoch 60	Epoch 110
mrcnn_bbox_loss	0.1936	0.1474
mrcnn_class_loss	0.1306	0.1450
mrcnn_mask_loss	0.3746	0.2676
rpn_bbox_loss	0.2347	0.2012

rpn_class_loss	0.1770	0.1504
loss	1.1100	0.9417

Η συνολική απόδοση(loss) στα δεδομένα εκπαίδευσης μετά την ολοκλήρωση και των 110 εποχών φτάνει το 0,9417. Αυτό αποτελεί το άθροισμα των επιμέρους δεικτών. Ξεχωριστά οι επιμέρους δείκτες, εκτός του mrcnn_mask_loss φαίνεται να έχουν ικανοποιητικές τιμές (<0,2). Ο mrcnn_mask_loss έχει τιμή 0,676 το οποίο επηρεάζει αρνητικά κατά 50% την συνολική απόδοση. Αυτό προκύπτει από το ότι το άθροισμα των επιμέρους δεικτών είναι περίπου 0,67 και προσθέτοντας την τιμή mrcnn_mask_loss προκύπτει το τελικό 0.9417. Ουσιαστικά ο δείκτης αυτός υποδεικνύει την ακρίβεια του μοντέλου στο να προβλέπει το που βρίσκονται βλάβες σε μια εικόνα δοντιού. Είναι ο ενδεχομένως ο δείκτης με την μεγαλύτερη βαρύτητα οπότε αυτό από μόνο τους προμηνύει ότι η απόδοση του μοντέλου δεν θα είναι ιδιαίτερα αυξημένη.



Εικόνα 13. Μετρικές απόδοσης στα δεδομένα επικύρωσης για την δοκιμή 1.

	Epoch 60	Epoch 110
mrcnn_bbox_loss	0.5246	0.4676
mrcnn_class_loss	0.5207	0.3384
mrcnn_mask_loss	0.5466	0.5923
rpn_bbox_loss	0.2901	0.2527
rpn_class_loss	0.1878	0.1771
loss	2.070	1.828

Στην συνέχεια οι ίδιοι δείκτες στα δεδομένα επικύρωσης έχουν μια διαφορετική εικόνα και δεν φαίνεται να έχουν πτωτική τάση καθώς προχωράνε οι εποχές. Και πάλι για τον δείκτη val_mrcnn_mask_loss φαίνεται ότι για τις πρώτες 40 εποχές υπάρχει πτώση στην τιμή του κάτι που υποδεικνύει ότι γίνεται πιο ακριβής, ενώ στην συνέχεια και ιδιαίτερα στις τελευταίες 30 εποχές η απόδοση πέφτει σημαντικά κάτι που υποδεικνύει overfitting του μοντέλου στα δεδομένα εκπαίδευσης για το συγκεκριμένο χαρακτηριστικό. Όσον αφορά τους υπόλοιπους δείκτες φαίνεται να βελτιώνονται συνεχώς, όμως μετά την 40^η εποχή υπάρχει κάμψη στην βελτίωση της απόδοσης.

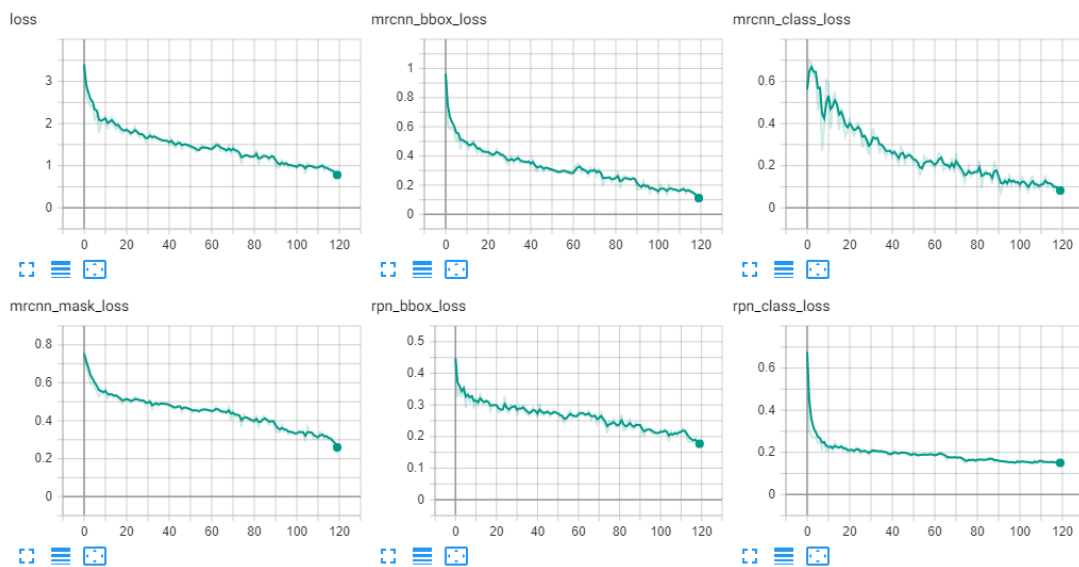
Το συγκεκριμένο πρόγραμμα εκπαίδευσης δημιούργησε ένα μοντέλο το οποίο έχει απόδοση 15.4%.

5.3.1.2 Δοκιμή 2

Η επόμενη δοκιμή που γίνεται, είναι για ένα πρόγραμμα εκπαίδευσης που ξεκινάει με 60 εποχές ενημερώνοντας μόνο τα τελευταία επίπεδα που αφορούν την κατηγοριοποίηση, και στην συνέχεια άλλες 40 εποχές ενημερώνοντας το νευρωνικό από το τέταρτο επίπεδο του Resnet και μετά. Αφού ολοκληρωθούν οι παραπάνω εποχές το πρόγραμμα ολοκληρώνεται με 20 ακόμα εποχές στις οποίες και πάλι ενημερώνεται όλο το δίκτυο. Ο ρυθμός μάθησης που χρησιμοποιήθηκε είναι και πάλι 0.001.

Πρόγραμμα εκπαίδευσης:

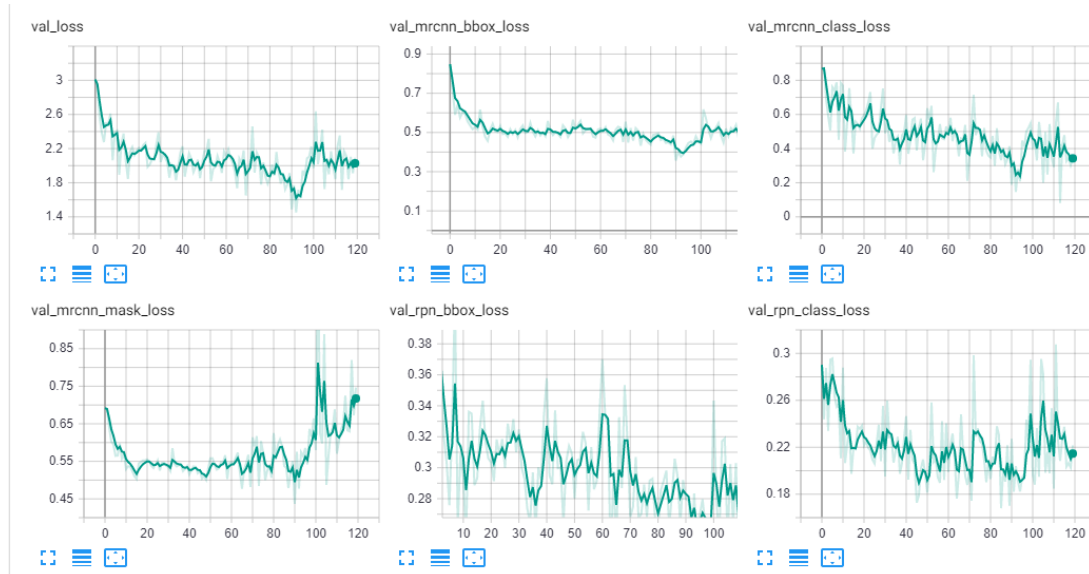
- i. 60 εποχές ενημερώνοντας μόνο τα βάρη από τελευταία επίπεδα (heads)
- ii. 50 εποχές ενημερώνοντας τα βάρη από το 4^ο επίπεδο του resnet και μετά (4+)
- iii. 20 εποχές ενημερώνοντας τα βάρη όλων των επιπέδων (all)



Εικόνα 14. Μετρικές απόδοσης στα δεδομένα εκπαίδευσης για την δοκιμή 2.

	Epoch 60	Epoch 100	Epoch 120
mrcnn_bbox_loss	0.2850	0.1405	0.0864
mrcnn_class_loss	0.1919	0.0939	0.0672
mrcnn_mask_loss	0.4466	0.3322	0.3376
rpn_bbox_loss	0.2794	0.2195	0.1725
rpn_class_loss	0.1831	0.1502	0.1519
loss	1.3860	0.9363	0.7047

Όπως και στο προηγούμενο πρόγραμμα υπάρχει ανάλογη συμπεριφορά από το νευρωνικό παρόλο που το πρόγραμμα διαφέρει αρκετά από τις 60 εποχές και μετά.



Εικόνα 15. Μετρικές απόδοσης στα δεδομένα επικύρωσης για την δοκιμή 2.

	Epoch 60	Epoch 100	Epoch 120
mrcnn_bbox_loss	0.5197	0.4426	0.4712
mrcnn_class_loss	0.4753	0.4080	0.3442
mrcnn_mask_loss	0.5474	0.5801	0.7464
rpn_bbox_loss	0.3699	0.3433	0.2945
rpn_class_loss	0.1990	0.1966	0.2216
loss	2.111	1.971	2.353

Και σε αυτή την δοκιμή η πρόοδος του δείκτη `val_mrcnn_mask_loss` υποδηλώνει overfitting στα δεδομένα εκπαίδευσης, ιδιαίτερα από την 90^η εποχή και μετά που παρατηρείται σχεδόν διπλασιασμός στην τιμή του κάτι που μεταφράζεται σε δύο φορές χειρότερη απόδοση στον εντοπισμό των αντικειμένων στις εικόνες εισόδου. Στις υπόλοιπες μετρικές οι τιμές μετά το πέρας των πρώτων εποχών δεν εμφανίζουν σημαντική βελτίωση.

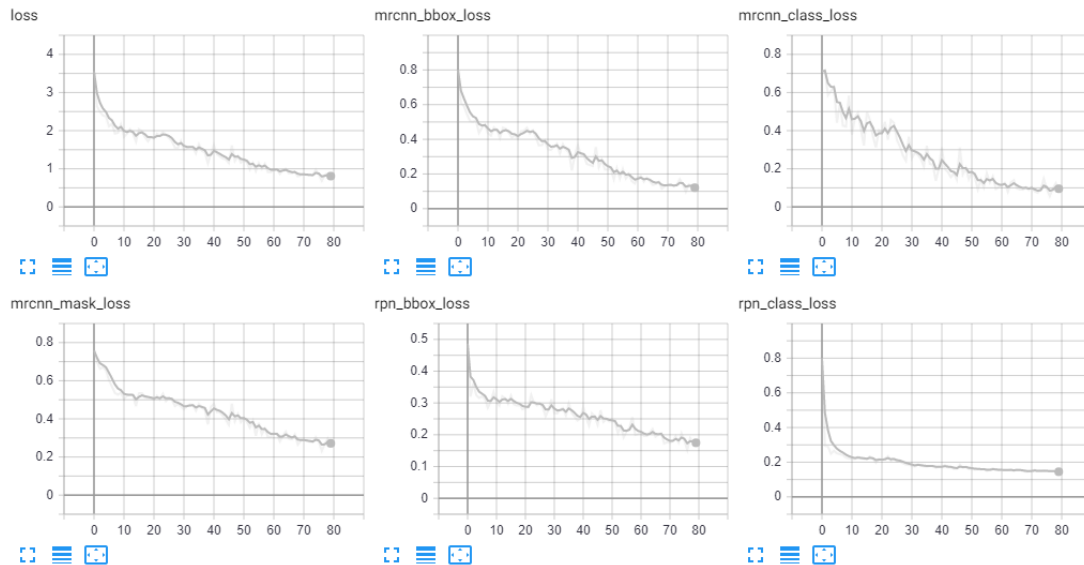
Ωστόσο το μοντέλο έχει μέση απόδοση 13.7% και όπως φαίνεται και στις παρακάτω εικόνες η απόδοση μπορεί να είναι χαμηλή όμως το μοντέλο καταφέρνει και εντοπίζει όλες τις περιοχές που έχουν βλάβες όμως όπως φαίνεται και από την τιμή του `mrcnn_class_loss` (~0.34) δεν είναι ακριβές στην επιλογή της κλάσης.

5.3.1.3 Δοκιμή 3

Στην συνέχεια το πρόγραμμα που δοκιμάστηκε αποτελείται από τρία στάδια. Στο πρώτο ενημερώνονται μόνο τα τελευταία στάδια του νευρωνικού δικτύου για 20 εποχές, μετά ακολουθεί ένα στάδιο στο οποίο για 40 εποχές ενημερώνονται τα επίπεδα του νευρωνικού που είναι μετά το τέταρτο επίπεδο του Resnet. Τέλος το πρόγραμμα ολοκληρώνεται με 40 εποχές σε όλο το δίκτυο για να προσαρμοστούν και τα υπόλοιπα επίπεδα. Σε αυτό το πρόγραμμα παρατηρείται ότι το μοντέλο αρχίζει να αποδίδει γρηγορότερα και καλύτερα στα δεδομένα εκπαίδευσης σε σχέση με τα προηγούμενα προγράμματα.

Πρόγραμμα εκπαίδευσης:

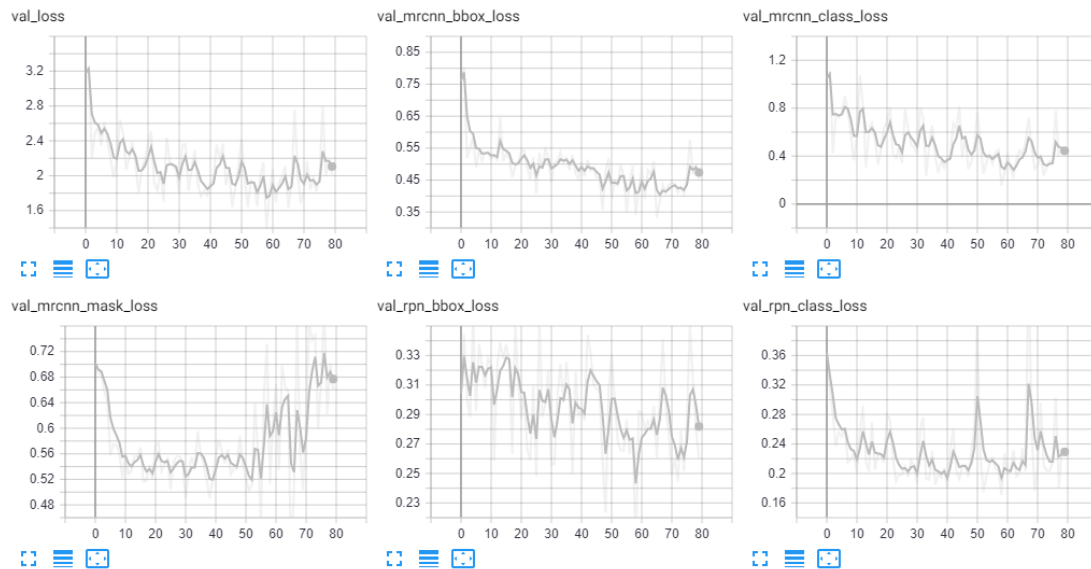
- i. 20 εποχές ενημερώνοντας μόνο τα βάρη από τελευταία επίπεδα (heads)
- ii. 20 εποχές ενημερώνοντας τα βάρη από το 4^ο επίπεδο του resnet και μετά (4+)
- iii. 40 εποχές ενημερώνοντας τα βάρη όλων των επιπέδων (all)



Εικόνα 16. Μετρικές απόδοσης στα δεδομένα εκπαίδευσης για την δοκιμή 3.

	Epoch 20	Epoch 40	Epoch 80
mrcnn_bbox_loss	0.4035	0.3661	0.1053
mrcnn_class_loss	0.3880	0.3129	0.0836
mrcnn_mask_loss	0.5007	0.4788	0.2607
rpn_bbox_loss	0.2854	0.2875	0.1684
rpn_class_loss	0.2152	0.1730	0.1415
loss	1.7930	1.6180	0.7595

Ενώ το συγκεκριμένο πρόγραμμα φαίνεται να αποδίδει καλύτερα από τα προηγούμενα στο σύνολο δεδομένων εκπαίδευσης, στα δεδομένα επικύρωσης φαίνεται να μην αποδίδει το ίδιο καλά. Πιο συγκεκριμένα το μοντέλο φαίνεται να κάνει overfit στο σύνολο δεδομένων που χρησιμοποιείται για εκπαίδευση. Αυτό παρατηρείται από τον δείκτη `val_mrcnn_mask_loss` ο οποίος φαίνεται να μεγαλώνει η τιμή του από την 35^η εποχή και μετά.



Εικόνα 17. Μετρικές απόδοσης στα δεδομένα επικύρωσης για την δοκιμή 3.

	Epoch 20	Epoch 40	Epoch 80
mrcnn_bbox_loss	0.5391	0.5084	0.4503
mrcnn_class_loss	0.6811	0.3856	0.4080
mrcnn_mask_loss	0.5740	0.5583	0.6599
rpn_bbox_loss	0.2875	0.2923	0.2601
rpn_class_loss	0.2796	0.1800	0.2345
loss	2.3610	1.9250	2.0130

Το παραπάνω μοντέλο έχει μέση απόδοση 12.25% σε νέα δεδομένα και είναι το μοντέλο με την μικρότερη απόδοση. Ωστόσο όπως φαίνεται και στις παρακάτω εικόνες το μοντέλο μπορεί και πάλι και δώσει μια κατεύθυνση σε κάποιον ειδικό για τις πιθανές βλάβες σε ένα δόντι.

5.3.1.4 Προβλέψεις σε εικόνα

Στην συνέχεια υπάρχει μια φωτογραφία με τους χαρακτηρισμούς που προστέθηκαν από τους γιατρούς και τις προβλέψεις που κάνει το μοντέλο με τα βάρη που έχουν προκύψει από κάθε πρόγραμμα εκπαίδευσης.



Εικόνα 18. Προτότυπη εικόνα με επισήμανση βλαβών.



Εικόνα 19. Προβλέψεις για βλάβες στην παραπάνω εικόνα από το μοντέλο που προέκυψε από το πρόγραμμα εκπαίδευσης της δοκιμής 1 (60 epochs head, 50 epochs all).



Εικόνα 20. Προβλέψεις για βλάβες στην παραπάνω εικόνα από το μοντέλο που προέκυψε από το πρόγραμμα εκπαίδευσης της δοκιμής 2 (60 epochs head, 40 epochs 4+, 20 epochs all).



Εικόνα 21. Προβλέψεις για βλάβες στην παραπάνω εικόνα από το μοντέλο που προέκυψε από το πρόγραμμα εκπαίδευσης της δοκιμής 3 (20 epochs head, 20 epochs 4+, 40 epochs all).

5.3.2 Δοκιμές στην προσαρμογή των προβλέψεων

Στην ενότητα αυτή γίνεται προσαρμογή των προβλέψεων του μοντέλου έτσι ώστε η πρόβλεψη να γενικευτεί από εικονοστοιχείο, σε υπερεικονοστοιχείου. Το μοντέλο επεξεργάζεται την εικόνα εισόδου και κάνει προβλέψεις σε περιοχές της εικόνας αυτής. Όμως κατά την εκπαίδευση οι εικόνες εισόδου έχουν μαρκαρισμένες περιοχές (υπερεικονοστοιχεία) οι οποίες αντιστοιχούν σε βλάβη συγκεκριμένης κλάσης. Οπότε στις παρακάτω δοκιμές επαναδημιουργούνται τα υπερεικονοστοιχεία τα οποία χρησιμοποιήθηκαν για τον χαρακτηρισμό και στην συνέχεια γενικεύοντας την πρόβλεψη του νευρωνικού, χαρακτηρίζονται τα υπερεικονοστοιχεία πλέον με κάποια κλάση.

Για κάθε ένα υπερεικονοστοιχείο εξετάζεται αν υπάρχουν εικονοστοιχεία που έχουν χαρακτηριστεί σαν βλάβες από το νευρωνικό και τότε υπολογίζεται η κλάση του υπερεικονοστοιχείου με βάση το πλήθος των εικονοστοιχείων που αντιστοιχούν σε κλάσεις. Οπότε το πλήθος των εικονοστοιχείων της ίδιας κλάσης διαμορφώνουν και την τελική κλάση που θα χαρακτηριστεί το υπερεικονοστοιχείο.

Για τις δοκιμές αυτές χρησιμοποιήθηκε το μοντέλο που προέκυψε από το πρώτο πρόγραμμα εκπαίδευσης, δηλαδή από την ενημέρωση για 60 εποχές τα τελικά επίπεδα του δικτύου και στην συνέχεια για ακόμα 50 εποχές ενημερώθηκε ολόκληρο το νευρωνικό. Το μοντέλο αυτό φάνηκε να αποδίδει καλύτερα από τα υπόλοιπα τα οποία δοκιμάστηκαν.

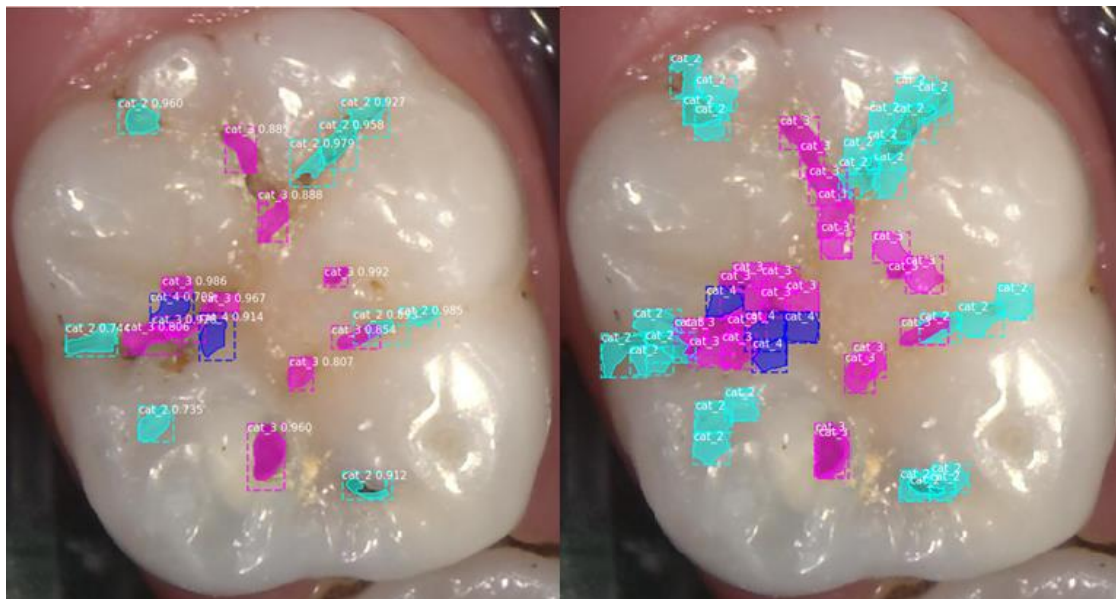
Μια καλύτερη όψη όσον αφορά τα αποτελεσματικότητα της υλοποίησης φαίνεται και αν δημιουργηθεί ένας πίνακας σύγχυσης για όλα τα εικονοστοιχεία. Με αυτόν τον τρόπο μπορεί εύκολα να παρατηρηθεί ότι ο αλγόριθμος σε κάποιες κλάσεις φαίνεται να είναι ακριβής, ενώ παράλληλα φαίνεται να υπάρχουν και αρκετές κακές προβλέψεις σε γειτονικές κλάσεις (κλάσεις 2, 3, 4). Αυτό είναι κάτι που παρατηρήθηκε και κατά την εκπαίδευση καθώς ο μόνος δείκτης που φάνηκε να μην βελτιώνεται σημαντικά ήταν ο δείκτης `mnem_mask_loss` ο οποίος αντιπροσωπεύει την ακρίβεια του μοντέλου στην κλάση.

5.3.2.1 Παράδειγμα 1

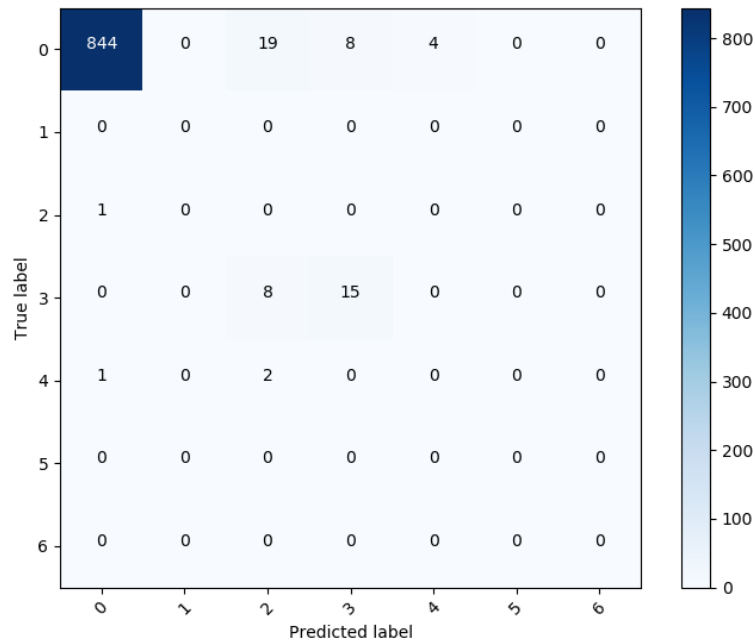
Στην παρακάτω εικόνα οι προβλέψεις που έχουν γίνει σε 894 υπερεικονοστοιχεία. Από αυτά οι 844 έχουν προβλεφθεί σωστά σαν περιοχές χωρίς βλάβη οπότε δεν έχουν ενδιαφέρον. Όσον αφορά τις υπόλοιπες 50 οι 15 έχουν προβλεφθεί σωστά. Από τις υπόλοιπες 35, οι 10 εξ αυτών εμφανίζουν ενδιαφέρον καθώς η πρόβλεψη έχει γίνει για γειτονικές κλάσεις και πιο συγκεκριμένα για τις κλάσεις 2, 3, 4.



Εικόνα 22 Εικόνα με χαρακτηρισμούς όπως δόθηκε από τους οδοντίατρος



Εικόνα 23. Παράδειγμα εικόνας δοντιού, αριστερά η εικόνα με τις προβλέψεις του νευρικού και δεξιά η εικόνα με τις γενικευμένες προβλέψεις σε επίπεδο υπερεικονοστοιχείων.



Εικόνα 24 Πίνακας σύγχυσης για το παράδειγμα 1 με γενικευμένες προβλέψεις.

Ακρίβεια	95.23%
Ισορροπημένη ακρίβεια	40.41%
Ακρίβεια (χωρίς σωστή πρόβλεψη για την κλάση 0)	25.86%
Ισορροπημένη ακρίβεια (χωρίς σωστή πρόβλεψη για την κλάση 0)	16.30%

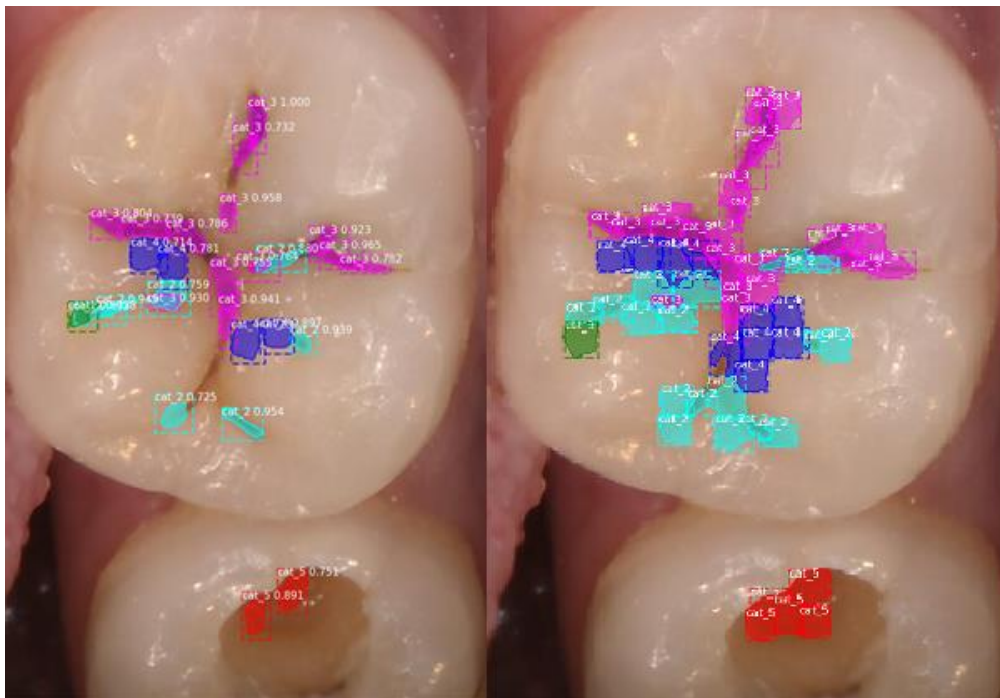
Πίνακας 3 μετρικές απόδοσης παραδείγματος 1

5.3.2.2 Παράδειγμα 2

Μια ακόμα εικόνα που συμπεριλαμβάνει δόντια με βλάβες από 4 διαφορετικές κλάσεις, το νευρωνικό φαίνεται να προσεγγίζει αρκετά καλά τις περιοχές με τις βλάβες όμως και πάλι σε αρκετές περιπτώσεις φαίνεται να μην προσεγγίζει σωστά την κλάση.

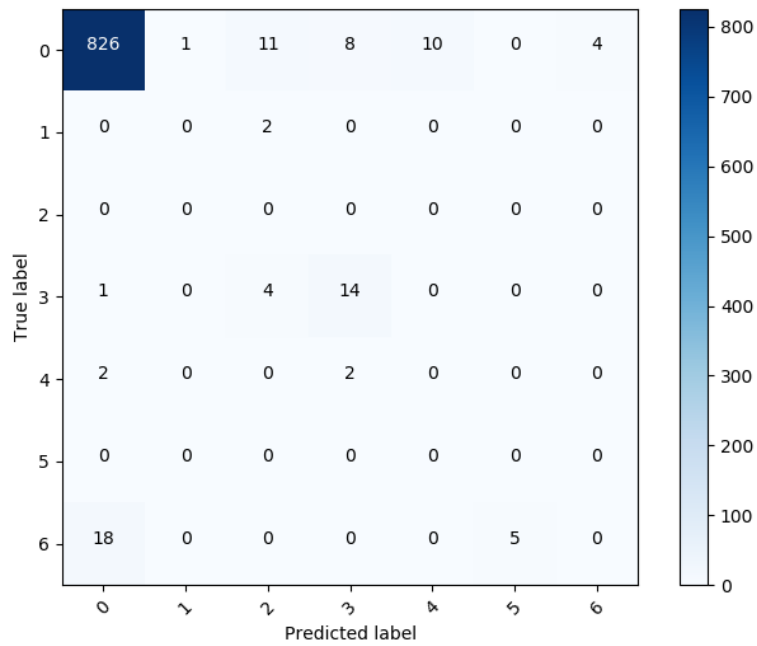


Εικόνα 25 Αρχική εικόνα με τους χαρακτηρισμούς.



Εικόνα 26 Αριστερά η εικόνα με τις προβλέψεις του νευρωνικού και δεξιά η εικόνα με τις γενικευμένες προβλέψεις σε επίπεδο υπερεικονοστοιχείων.

Και σε αυτή την περίπτωση προβλέψεις για τις κλάσεις 2, 3 έχουν προβλεφθεί λάθος μεταξύ τους.

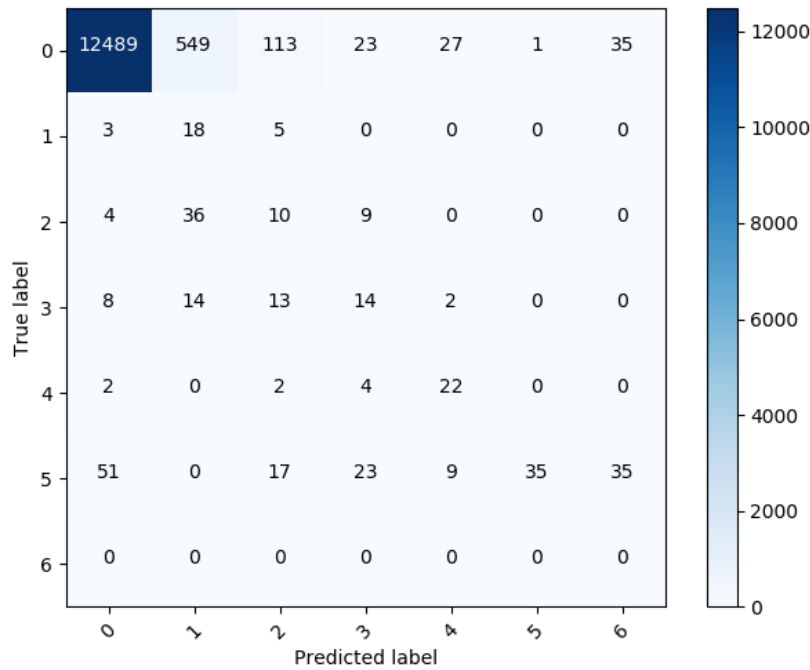


Εικόνα 27 Πίνακας σύγχυσης για το παράδειγμα 2 με γενικευμένες προβλέψεις.

Ακρίβεια	93.2%
Ισορροπημένη ακρίβεια	72.06%
Ακρίβεια (χωρίς σωστή πρόβλεψη για την κλάση 0)	25.61%
Ισορροπημένη ακρίβεια (χωρίς σωστή πρόβλεψη για την κλάση 0)	48.05%

Πίνακας 4 μετρικές απόδοσης παραδείγματος 2.

Συνολικά για όλες τις εικόνες επικύρωσης και το συγκεκριμένο μοντέλο με τις προσαρμογές στις προβλέψεις, δημιουργείται ο παρακάτω συγκεντρωτικός πίνακας σύγχυσης. Παρακάτω υπάρχουν και οι μετρικές που συμπεριλαμβάνουν όλες τις προβλέψεις αλλά και αυτές που δεν συμπεριλαμβάνουν τις σωστές προβλέψεις για το την κλάση 0.



Εικόνα 28 Πίνακας σύγχυσης απο την συνολική απόδοση στις εικόνες επικύρωσης.

Ακρίβεια	92.74%
Ισορροπημένη ακρίβεια	50.31%
Ακρίβεια (χωρίς σωστή πρόβλεψη για την κλάση 0)	9.13%
Ισορροπημένη ακρίβεια (χωρίς σωστή πρόβλεψη για την κλάση 0)	34.59%

Πίνακας 5 μετρικές συνολικής απόδοσης στο σύνολο των εικόνων επικύρωσης

5.3.3 Δοκιμές σε προσαρμοσμένες κλάσεις

Προσπαθώντας να αντιμετωπίσουμε το πρόβλημα με τις γειτονικές κλάσεις απευθυνθήκαμε στους γιατρούς για να ερμηνεύσουν ιατρικά την ομοιότητα σε γειτονικές κλάσεις. Οι κλάσεις που χρησιμοποιήθηκαν στηρίζονται στο σύστημα ταξινόμησης ICDAS-II το οποίο δημιουργήθηκε αφενός για να κατατάξει τις τερηδονικές βλάβες σύμφωνα με τα διάφορα στάδια εξέλιξης τους με βάση οπτικά κριτήρια αλλά και κυρίως με βάση αυτήν την κατάταξη να μπορεί να προσδιορίσει την προτεινόμενη θεραπεία για την κάθε βλάβη.

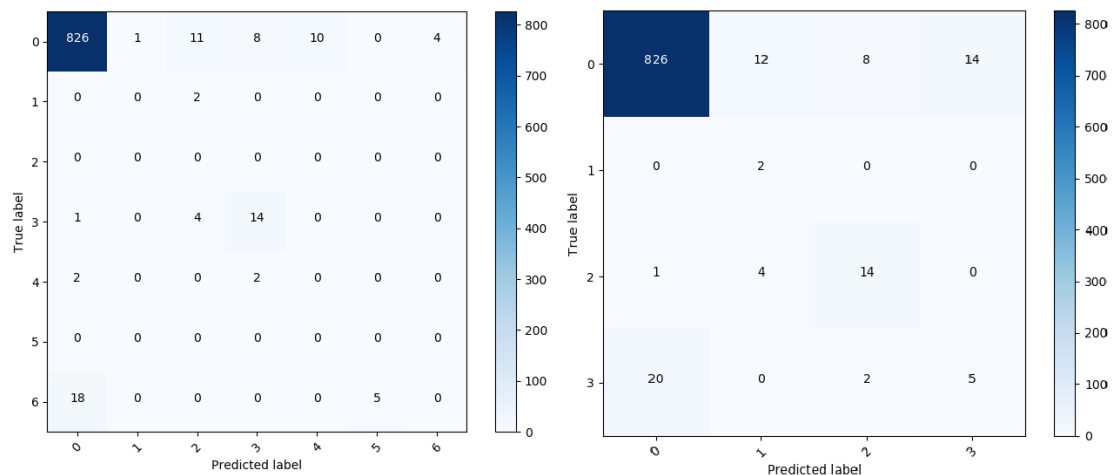
Από κλινική άποψη επομένως θα ήταν σημαντικό τα ταξινομηθούν οι βλάβες σε 4 κατηγορίες ανάλογα με την ανάγκη θεραπείας που έχουν. Οι τέσσερις αυτές κατηγορίες θα ήταν υγιείς επιφάνειες (A), αρχόμενες βλάβες που απαιτούν προληπτική παρέμβαση (B), επιφάνειες με οριακή απώλεια αδαμαντίνης (C) και μέτριες ή εκτεταμένες βλάβες που χρειάζονται αποκατάσταση (D). Με αυτό το σκεπτικό οι 7 (6 κλάσεις με βλάβες και η υγιής κλάση ή φόντο στην περίπτωση μας) κατηγορίες του συστήματος ICDAS II μπορούν να ενωθούν σε τρεις κατηγορίες όπως φαίνεται στον παρακάτω πίνακα.

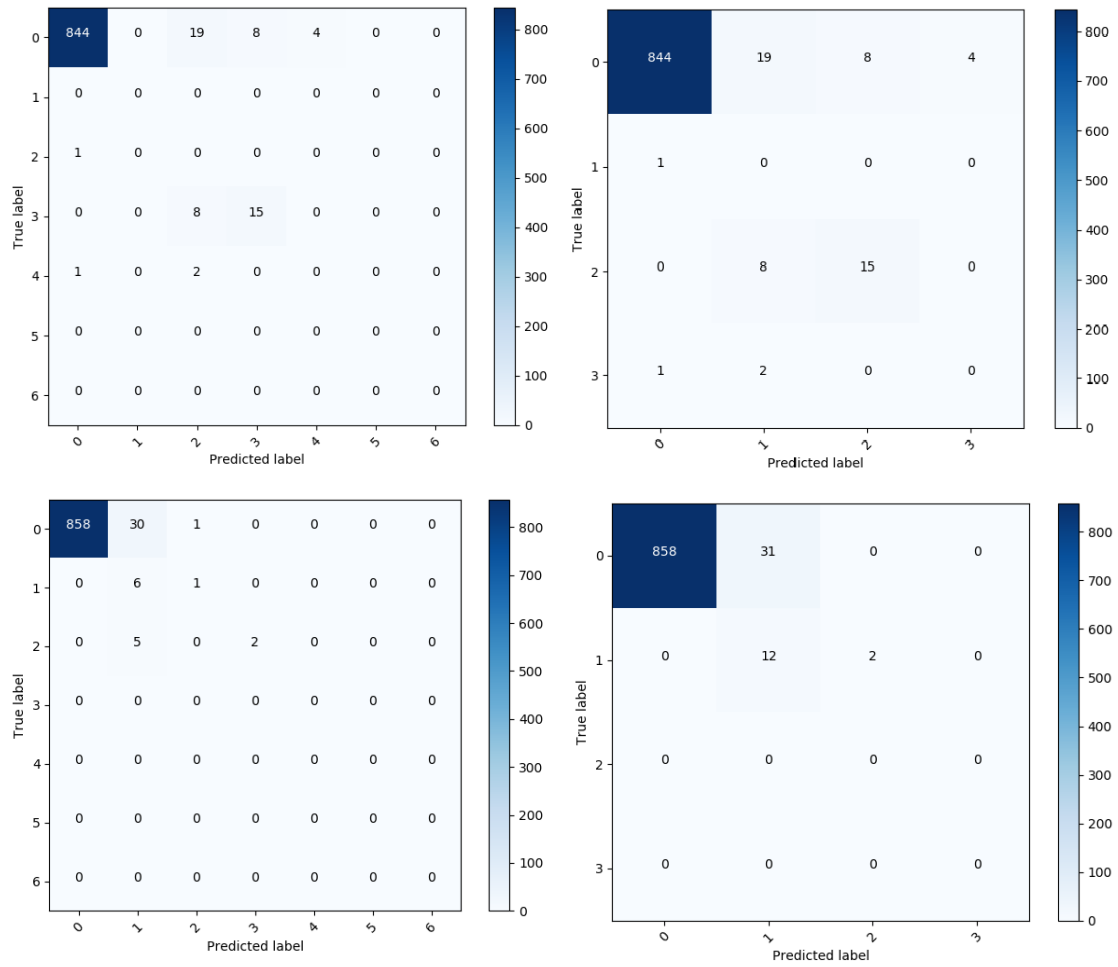
Αυτές οι κατηγορίες θα έχουν σαφώς μεγαλύτερη κλινική σημασία και είναι οι παρακάτω:

Κατηγορία ICDAS	Νέα κατηγοριοποίηση
0	A (υγιής επιφάνεια)
1	B
2	(αρχόμενες βλάβες)
3	C
4	Οριακή απώλεια αδαμαντίνης
5	D
6	(μέτριες και εκτεταμένες βλάβες)

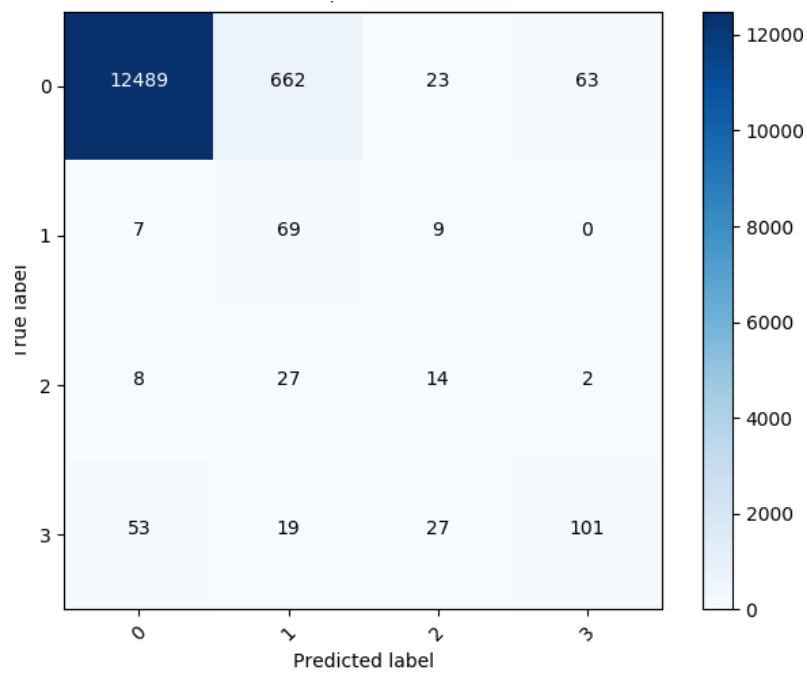
Η κατηγορία A αφορά μια υγιή περιοχή του δοντιού. Η κατηγορία B αφορά περιοχές του δοντιού που αρχίζουν να εμφανίζουν μια απόκλιση από την φυσιολογική δομή του δοντιού και χρειάζονται προληπτική παρέμβαση. Η κατηγορία C αφορά τις πρώτες μεταβολές στην επιφάνεια του δοντιού που θα έπρεπε να αντιμετωπιστούν με την παρασκευή (τροχισμό) της επιφάνειας. Τέλος, οι επιφάνειες που ανήκουν στην κατηγορία D είναι περιοχές του δοντιού που έχουν σημαντική βλάβη και πρέπει να τροχιστούν (να «πεθάνουν») ώστε να σταματήσει η συνέχιση της καταστροφής του δοντιού.

Οπότε για τις δοκιμές στην απόδοση της υλοποίησης με τις ενοποιημένες κατηγορίες, χρησιμοποιήθηκε το ίδιο μοντέλο με την προηγούμενη σειρά δοκιμών, το οποίο έχει προκύψει από το πρόγραμμα εκπαίδευσης (60 εποχές heads, 50 all). Αφού το μοντέλο έκανε κάποιες προβλέψεις στην συνέχεια γενικεύθηκαν σε επίπεδο υπερεικονοστοιχείων (βλ. προηγούμενη σειρά δοκιμών). Τέλος οι προβλέψεις σε επίπεδο υπερεικονοστοιχείων είχαν γίνει στην κλίμακα ICDAS-II οπότε υπήρχαν 6 (+1 το φόντο) κλάσεις. Χρησιμοποιώντας την παραπάνω αντιστοίχιση των κατηγοριών δημιουργούνται 3 (+1 το φόντο) κλάσεις πλέον. Οπότε αξιοποιώντας τις προσαρμοσμένες προβλέψεις του νευρωνικού σε επίπεδο υπερεικονοστοιχείων, ενοποιώντας τις κλάσεις σε νέες κλινικά πιο ακριβείς προκύπτουν οι παρακάτω πίνακες σύγχυσης.





Εικόνα 29 αριστερά οι πίνακες σύγκρισης από γενικευμένες προβλέψεις σε 6 (+1) κλάσεις και δεξιά οι πίνακες σύγκρισης από τις ίδιες προβλέψεις σε ενοποιημένες 3 (+1) κλάσεις.



Εικόνα 30 Πίνακας σύγκρισης για το σύνολο των εικόνων επικύρωσης με τις νέες ενοποιημένες κλάσεις.

Ακρίβεια	93.37%
Ισορροπημένη ακρίβεια	63.37%
Ακρίβεια (χωρίς σωστή πρόβλεψη για την κλάση 0)	16.97%
Ισορροπημένη ακρίβεια (χωρίς σωστή πρόβλεψη για την κλάση 0)	39.78%

Πίνακας 6 μετρικές συνολικής απόδοσης στις εικόνες του συνόλου επικύρωσης

Η ακρίβεια του αλγόριθμου φαίνεται τελικά να βελτιώνεται ακόμα περισσότερο όταν οι προβλέψεις προσαρμόζονται στην νέα κλίμακα κλάσεων. Εδώ πλέον, όπως φαίνεται και από την διαγώνιο των παραπάνω πινάκων, υπάρχει σημαντική βελτίωση στην ακρίβεια αφού πλέον έχουμε ανάγκη τις κλάσεις σε νέες κλινικά παρόμοιες.

6 Συμπεράσματα

Αξιολογώντας τα αποτελέσματα της πειραματικής διαδικασίας οδηγούμαστε στο συμπέρασμα ότι ο εντοπισμός βλαβών σε ενδοστόμιες εικόνες δοντιών είναι αρκετά σύνθετος.

Το Mask RCNN είναι ένα νευρωνικό το οποίο μπορεί να γενικευτεί και να προσαρμοστεί ακόμα και σε εικόνες με περιοχές που δεν είναι τόσο ευδιάκριτες μεταξύ τους. Είναι σχετικά αποτελεσματικό στο κομμάτι εντοπισμού μιας περιοχής η οποία αντιπροσωπεύει μια βλάβη, όμως δυσκολεύεται αρκετά στον εντοπισμό της κλάσης μιας περιοχής. Οι κλάσεις οι οποίες φαίνεται να μην εντοπίζονται με τόση ακρίβεια είναι κυρίως οι πιο συνηθισμένες (2, 3, 4) οι οποίες έχουν και παρόμοια κλινικά χαρακτηριστικά. Κάτι το οποίο παρατηρήθηκε από τις μετρικές απόδοσης του νευρωνικού και πιο συγκεκριμένα από την μετρική `mrcnn_mask_loss` η οποία δεν βελτιωνόταν με τον ρυθμό που βελτιωνόντουσαν οι υπόλοιπες μετρικές και μάλιστα από κάποια εποχή και μετά ξεκινούσε να κάνει `overfit` στο σύνολο εκπαίδευσης.

Η κλίμακα ICDAS-II που χρησιμοποιήθηκε για τον επιστημονικό χαρακτηρισμό των βλαβών, είναι μια κλίμακα οι οποία χρησιμοποιείται για να χαρακτηρίσει οδοντικές περιοχές στηριζόμενη κατά κύριο λόγο σε οπτικά χαρακτηριστικά της εκάστοτε περιοχής.

Χρησιμοποιώντας την νέα κλίμακα που προτάθηκε από τους γιατρούς που η νέα ενοποιημένη πλέον κλίμακα έδωσε πολύ καλύτερα αποτελέσματα μιας και διπλασιάστηκε σχεδόν ο δείκτης της ακρίβειας, ενώ ο δείκτης της ισορροπημένης ακρίβειας αυξήθηκε αρκετά επίσης.

	ICDAS-II	Νέα κλίμακα	Διαφορά
Ακρίβεια	92.74%	93.37%	+0.6%
Ισορροπημένη ακρίβεια	50.31%	63.37%	+25.9%
Ακρίβεια (χωρίς σωστή πρόβλεψη για την κλάση 0)	9.13%	16.97%	+85%
Ισορροπημένη ακρίβεια (χωρίς σωστή πρόβλεψη για την κλάση 0)	34.59%	39.78%	+15%

Πίνακας 7 σύγκριση μετρικών απόδοσης ανάμεσα στην κλίμακα ICDAS-II και την νέα κλίμακα.

Επίσης είναι πολύ σημαντικό να αναφερθεί ότι και το περιορισμένο σύνολο δεδομένο ήταν ένας ανασταλτικός παράγοντας στην αύξηση της ακρίβειας των προβλέψεων του νευρωνικού δικτύου. Αυτό οφείλεται στο ότι η λήψη σε πρώτη φάση των φωτογραφιών και στην συνέχεια η διαδικασία χαρακτηρισμού αυτών με τις βλάβες που υπάρχουν, είναι από μόνη της μια διαδικασία που απαιτεί αρκετό χρόνο για την κάθε εικόνα ξεχωριστά.

Εν κατακλείδι τα νευρωνικά δίκτυα, τα οποία αναπτύχθηκαν για να αντιμετωπίσουν προβλήματα επεξεργασίας δεδομένων εικόνας και ήχου, μπορούν να χρησιμοποιηθούν και σε τομείς όπως η ιατρική δίνοντας επιπλέον εργαλειακές επιλογές για πιο ακριβής διαγνώσεις ακόμα και απομακρυσμένα σε περιοχές που δεν είναι εύκολη η πρόσβαση πάντα.

7 Ευχαριστίες

Θα ήθελα να ευχαριστήσω τον καθηγητή Ηλία Μαγκλογιάννη του τμήματος Ψηφιακών συστημάτων του Πανεπιστημίου Πειραιώς για την επίβλεψη σε αυτή την έρευνα καθώς και για την πολύτιμη βοήθεια και καθοδήγηση που μου παρείχε προκειμένου να ολοκληρώσω αυτήν την διατριβή.

Ακόμα θα ήθελα να αναφέρω ότι χωρίς την συνεισφορά και την πολύτιμη βοήθεια των γιατρών, η συγκεκριμένη διατριβή δεν θα μπορούσε να εκπονηθεί. Πιο συγκεκριμένα θα ήθελα να ευχαριστήσω τους αξιότιμους γιατρούς δρ. Κωνσταντίνο Ουλή, δρ. Ηλία Μπερδούση καθώς και την δρ. Φανή Ανδρίκουλα για την βοήθεια τους προκειμένου να φωτογραφηθούν σωστά τα δόντια καθώς και τον χαρακτηρισμό αρκετών δοντιών έτσι ώστε να δημιουργηθούν δεδομένα αρκετά προκειμένου να φτάσουμε σε κάποια συμπεράσματα.

Τέλος θα ήθελα να ευχαριστήσω και τον κύριο Κωνσταντίνο Μούτσελο για την πολύτιμη καθοδήγηση, για την προετοιμασία που είχε γίνει ήδη καθώς και για το εργαλείο χαρακτηρισμού εικόνων που ανέπτυξε και βοήθησε στην δημιουργία των δεδομένων.

8 Παραρτήματα

8.1 Παράρτημα I. Dockerfile

Το Dockerfile το οποίο δημιουργεί το περιβάλλον για να τρέξει η εφαρμογή.

```
FROM tensorflow/tensorflow:1.13.1-gpu-py3
WORKDIR /var/project/data/annotator
COPY ./data/annotator /var/project/data/annotator
WORKDIR /var/project/src
COPY ./src /var/project/src
WORKDIR /var/project/data/output/training
WORKDIR /var/project/data/output/validation
WORKDIR /var/project/data/output/unknown
COPY requirements.txt /var/project/
WORKDIR /var/project
RUN apt-get update && apt-get install -y git
RUN pip install Cython
RUN pip install --requirement requirements.txt
RUN apt-get install -y libsm6 libxext6 libxrender-dev
CMD python -c "import tensorflow as tf; tf.enable_eager_execution();
print(tf.reduce_sum(tf.random_normal([1000, 1000])))"
```

8.2 Παράρτημα II. Εντολές εκκίνησης εφαρμογής

Εντολές εκκίνησης εφαρμογής από γραμμή εντολών.

Προαιρετικά ορίσματα:

```
-h, --help show this help message and exit
--model_dir MODEL_DIR Path to weights .h5 file or 'coco'
--init_with INIT_WITH coco, or last
--input_training_data INPUT_TRAINING_DATA input training data folder
--input_validation_data INPUT_VALIDATION_DATA input validation data folder
--input_unknown_data INPUT_UNKNOWN_DATA input unknown data folder
--training_data TRAINING_DATA final training data folder
--validation_data VALIDATION_DATA final validation data folder
```

--unknown_data UNKNOWN_DATA final unknown data folder

8.3 Παράρτημα III. Πρόγραμμα εκπαίδευσης

Βήματα του προγράμματος εκπαίδευσης που δημιουργούν τα βάρη με την μεγαλύτερη απόδοση.

```
model.train(data_train, data_val,  
            learning_rate=cfg.LEARNING_RATE,  
            epochs=60,  
            layers='heads',  
            augmentation=augmentation)  
  
model.train(data_train, data_val,  
            learning_rate=cfg.LEARNING_RATE,  
            epochs=110,  
            layers='all',  
            augmentation=augmentation)
```

8.4 Παράρτημα IV. Βοηθητικές βιβλιοθήκες προγραμματισμού σε Python

Βιβλιοθήκες που χρησιμοποιήθηκαν για να δημιουργηθεί η παραπάνω εφαρμογή είναι οι παρακάτω:

absl-py, astor, backcall, certifi, colorama, cycler, Cython, decorator, gast, grpcio, h5py, imageio, imgaug, ipython, ipython-genutils, jedi, Keras, Keras-Applications, Keras-Preprocessing, kiwisolver, Markdown, matplotlib, mkl-fft, mkl-random, mock, networkx, numpy, opencv-python, parso, pbr, pickleshare, Pillow, prompt-toolkit, protobuf, pycocotools, Pygments, pyparsing, pyreadline, python-dateutil, PyWavelets, PyYAML, read-roi, scikit-image, scipy, Shapely, six, tensorboard, tensorflow-gpu, tensorflow-estimator, termcolor, -e git+https://github.com/matterport/Mask_RCNN#egg=mrcnn, traitlets, wcwidth, Werkzeug, wincertstore.

9 Αναφορές

- [1] T.-Y. Lin, M. Maire, S. Belongie, L. Bourdev, R. Girshick, J. Hays, P. Perona, D. Ramanan, C. L. Zitnick και P. Dollár, «Microsoft COCO: Common Objects in Context,» 2014.
- [2] P. Sodhi, A. Sodhi, N. Vishal και S. Vishal, «Introduction to Machine Learning and Its Basic Application in Python,» σε *Proceedings of 10th International Conference on Digital Strategies for Organizational Success*, 2019.
- [3] I. Goodfellow, Y. Bengio και A. Courville, *The MIT Press*, 2016.
- [4] K. Suzuki, «Pixel-based machine learning in medical imaging,» *International Journal of Biomedical Imaging*, 2012.
- [5] I. Kononenko, «Machine learning for medical diagnosis: history, state of the art and perspective,» *Artificial Intelligence in Medicine*, 2001.
- [6] C. Nello, *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*, Cambridge University Press, 2013.
- [7] J. R. Quinlan, «Induction of decision trees,» 1986.
- [8] S. George και A. J. Lee, *Linear Regression Analysis*, 1989.
- [9] D. W. Hosmer και S. Lemeshow, *Applied Logistic Regression*, Second Edition, 2000.
- [10] D. Lowd και P. Domingos, «Naive Bayes models for probability estimation,» σε *International Conference Proceeding Series (ICP)*, 2005.
- [11] C. Yin Zhou και Y. QiuChen, «Improving nearest neighbor classification with cam weighted distance,» 2006.
- [12] L. Breiman, «Random Forests,» 2001.
- [13] K. Hornik, M. Stinchcombe και H. White, *Multilayer feedforward networks are universal approximators*, 1989.
- [14] K. Krishna και M. N. Murty, «Genetic K-means algorithm,» 1999.
- [15] D. Comaniciu και P. Meer, «Mean shift: a robust approach toward feature space analysis,» 2002.
- [16] D. Dueck και B. J. Frey, «Non-metric affinity propagation for unsupervised image categorization,» 2007.
- [17] S. C. Johnson, «Hierarchical clustering schemes,» 1967.
- [18] D. Birant και A. Kut, *ST-DBSCAN: An algorithm for clustering spatial–temporal data*, 2007.

- [19] S. Roberts, D. Husmeier, I. Rezek και W. Penny, «Bayesian approaches to Gaussian mixture modeling,» 1998.
- [20] S. Z.Li, Ch. 13. Modeling image analysis problems using markov random fields, 2003.
- [21] J. C. Dunn, «A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters,» 1973.
- [22] C. Bezdek, R. Ehrlich και W. Full, FCM: The fuzzy c-means clustering algorithm, 1984.
- [23] M. Nixon, M. Nixon και A. Aguado, «Feature extraction & image processing for computer vision,» *Academic Press*, 2012.
- [24] T. M. Mitchell, «Machine Learning,» σε *Machine Learning*, 1997.
- [25] T. A. M. S. A.-Z. Saad Albawi, «Understanding of a convolutional neural network,» σε *2017 International Conference on Engineering and Technology (ICET)*, 2017.
- [26] R. Yamashita, M. Nishio, R. Kinh Gian Do και K. Togashi, «Convolutional neural networks: an overview and application in radiology,» σε *Insights into Imaging Vol. 9*, 2018.
- [27] J. Gu, Z. Wang, J. Kuen και L. Ma, «Recent Advances in Convolutional Neural Networks,» 2018.
- [28] S. Ren, K. He, R. Girshick και J. Sun, «Faster R-CNN: Towards real-time object detection with region proposal network,» *In NIPS*, 2015.
- [29] G. Ross, «Fast R-CNN,» *In ICCV*, 2015.
- [30] P. O. Pinheiro, R. R. Collobert και P. Dollar, «Learning to segment object candidates,» *In NIPS*, 2015.
- [31] J. Dai, K. He και J. Sun, «Instance-aware semantic segmentation via multi-task network cascades,» *In CVPR*, 2016.
- [32] Y. Li, H. Qi, J. Dai, X. Ji και Y. Wei, «Fully convolutional,» *In CVPR*, 2017.
- [33] R. Girshick, J. Donahue, T. Darrell και J. Malik, «Rich feature hierarchies for accurate object detection and semantic segmentation,» *In CVPR*, 2014.
- [34] J. Long, E. Shelhamer και T. Darrell, «Fully convolutional networks for semantic segmentation,» *In CVPR*, 2015.
- [35] T.-Y. L. R. C. a. P. D. P. O. Pinheiro, «Learning to refine object segments,» *In ECCV*, 2016.
- [36] M. Jaderberg, K. Simonyan, A. Zisserman και K. Ka, «Spatial transformer networks,» *In NIPS*, 2015.
- [37] K. HE, X. Zhang, S. Ren και J. Sun, «Deep residual learning for image recognition,» *In CVPR*, 2016.

- [38] S. Xie, R. Girshick, P. Dollar, Z. Tu και K. He, «Aggregated residual transformations for deep neural networks,» *In CVPR*, 2017.
- [39] J. Huang, V. Rathod, C. Sun, M. Zhu και A. Korattikar, «Speed/accuracy trade-offs for modern convolutional object detectors,» *In CVPR*, 2017.
- [40] A. Shrivastava, R. Sukthankar, J. Malik και A. Gupta, «Beyond skip connections: Top-down modulation for object detection,» 2016.
- [41] T.-Y. Lin, P. Dollar, R. Girshick, K. He, B. Hariharan και S. Belongie, «Feature pyramid networks for object detection,» *In CVPR*, 2017.
- [42] K. He, G. Gkioxari, P. Dollár και G. Ross, «Mask R-CNN,» 2017.
- [43] A. Shrivastava, G. A και G. R., «Training regionbased object detectors with online hard example mining,» *In CVPR*, 2016.
- [44] A. Radhakrishna, S. Appu, S. Kevin, L. Aurelien, P. Fua και S. Sússtrunk, «SLIC Superpixels Compared to State-of-the-Art Superpixel Methods,» 2012.
- [45] A. Jablonski-Momeni, V. Stachniss και D. Ricketts, «Reproducibility and Accuracy of the ICDAS-II for Detection of Occlusal Caries in vitro,» 2008.
- [47] B. J. Erickson, P. Korfiatis, Z. Akkus and T. L. Kline, "Machine Learning for Medical Imaging," *The Radiological Society of North America*, p. 11, 2017.