ΕΡΓΑΣΙΑ ΣΤΟ DEEP LEARNING

Καλογερόπουλος Δημήτρης

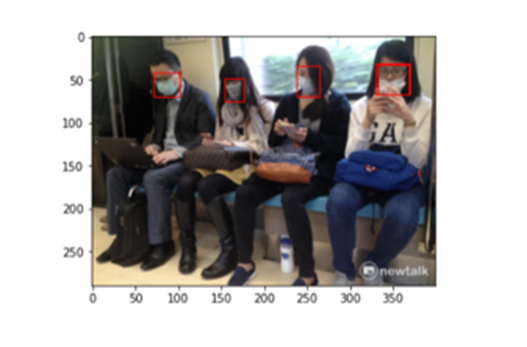
Mtn2202

**Eισαγωγή**

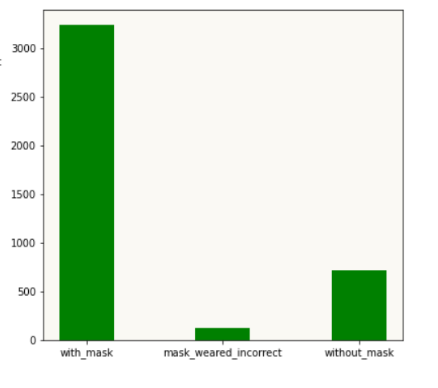
Η παρούσα εργασία ασχολείται με τον εντοπισμό μασκών σε τυχαίες φωτογραφίες ανθρώπων. Πιο συγκεκριμένα ελέγχεται η σωστή χρήση μάσκας ,η λανθασμένη χρήση της ( πχ κάτω από τη μύτη του ατόμου) καθώς και η καθόλου χρήση της σε φωτογραφίες οι οποίες περιέχουν διάφορα άτομα τα οποία δεν έχουν κάνει απαραίτητα την ίδια επιλογή εφαρμογής της. Πέραν από τη διαδικασία ταξινόμησης στις 3 αυτές κατηγορίες ένας δεύτερος σκοπός του μοντέλου είναι να μπορέσει να αποφανθεί στις διάφορες φωτογραφίες το που υπάρχουν πρόσωπα βάζοντας ένα κατάλληλο πλαίσιο αποφανθεί την κλάση που ανήκουν. Το μοντέλο που χρησιμοποιήθηκε είναι το FasteRCNN.

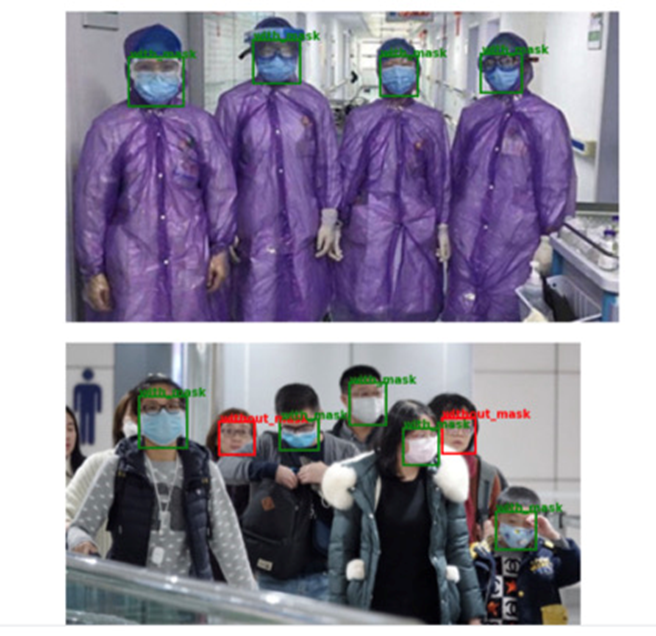
**Dataset-xml**

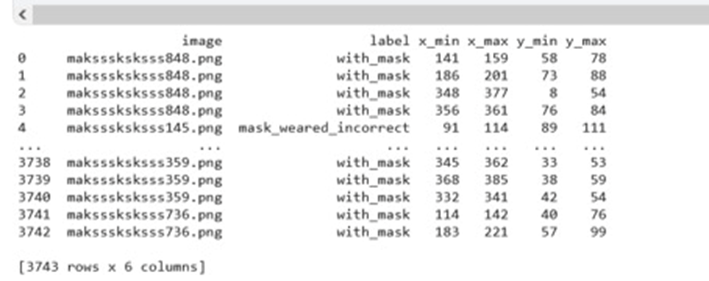
To dataset αποτελείται από 853 εικόνες οι οποίες έχουν συγκεκριμένο frame στα πρόσωπα και το αντίστοιχο labelling (mask, without\_mask, mask\_weared\_incorrectly). Σε αυτές το σύνολο των annotations είναι 3743 εκ των οποίων τα ποσοστά φαίνονται στα παρακάτω γραφήματα. Επιπλέον παρουσιάζεται ένα δείγμα των εικόνων. Τέλος πρέπει να τονιστεί πως τα annotations έχουν γίνει με τη χρήση xml ενώ οι φωτογραφίες σε png.



**PREPROCESSING**

Aρχικά συλλέγονται οι πληροφορίες σχετικά με το αν μια εικόνα περιέχει πρόσωπα με μάσκες ή χωρίς μάσκες ή με μάσκα που εφαρμόζεται λάθος Εξάγοντας και αποθηκεύοντας αυτές τις ετικέτες (π.χ. "with\_mask" ή "without\_mask"). Έτσι για να γίνει βέβαιο ότι τα στοιχειά έχουν εξαχθεί και συνδεθεί σωστά , γίνεται οπτικοποίηση των annotations στις αντίστοιχες εικόνες καθώς και χρήση αντίστοιχων χρωμάτων με βάση τα labels ( λευκό στην πρώτη περίπτωση ,μαύρο στη δεύτερη και κίτρινο στην τρίτη) .Στη συνέχεια μετατρέπονται σε dataframe οι εικόνες με τα αντίστοιχα labellings η κάθε μια τα οποία εν συνεχεία μετατράπηκαν σε 0,1,2 αντίστοιχα ώστε να μπορούν να επεξεργαστούν από το εκάστοτε μοντελό .Παρότι τα data ήταν imbalanced  με το augmentation που εφαρμόστηκε τα αποτελέσματα ήταν πολύ χειρότερα οπότε αυτό εγκαταλείφθηκε (κυρίως όσον αφορά το localation loss).Τέλος μέσω της κλάσης Μaskdataset δημιουργείται το dataset που περιέχει τα annotations με τiς αντίστοιχες κλάσεις ώστε να μπορέσει να αξιοποιηθεί προς το ζητούμενο task.Έτσι εξασφαλίζεται ότι το μοντέλο θα λάβει τόσο τα δεδομένα εικόνας όσο και το ground truth κατά τη διάρκεια της εκπαίδευσης σε κατάλληλη μορφή.





**MODEL**

Αρχικοποιείται το προεκπαιδευμένο μοντέλο Faster R-CNN που έχει ως backbone την αρχιτεκτονική ResNet-50 . Αυτό το προ-εκπαιδευμένο μοντέλο έχει ήδη εκπαιδευτεί σε ένα μεγάλο σύνολο δεδομένων δεδομένων COCO (Common Objects in Context) με στόχο το Object detection. Αντικαθιστά την υπάρχουσα κεφαλή ταξινομητή στο προ-εκπαιδευμένο μοντέλο με μια νέα (FastRCNNPredictor) που είναι προσαρμοσμένη στον αριθμό κλάσεων (3) που αφορούν το συγκεκριμένο task καθώς και 1 ακόμα που αφορά πάντα το background των φωτογραφιων. Αυτό το βήμα είναι απαραίτητο επειδή το αρχικό προ-εκπαιδευμένο μοντέλο έχει μια κεφαλή ταξινομητή που έχει σχεδιαστεί για διαφορετικό αριθμό κλάσεων (και πρέπει να προσαρμοστεί.Όσον αφορά το ReSnet -50 ,πρόκειται για μια αρχιτεκτονική συνελκτικου νευρωνικού δικτύου με skip connections οι οποιες επιτρέπουν στο δίκτυο να συλλαμβάνει χαρακτηριστικά σε διαφορετικές κλίμακες όπως πρόσωπα με ή χωρίς μάσκες ανεξάρτητα του που βρισκόνται στη φωτογραφία. Έτσι ολοκληρώνεται η προετοιμασία του μοντέλου.

Συνοπτικά, η συνάρτηση αυτή παρέχει έναν βολικό τρόπο φόρτωσης ενός προ-εκπαιδευμένου μοντέλου Faster R-CNN και προσαρμογής του για μια συγκεκριμένη εργασία ανίχνευσης αντικειμένων με διαφορετικό αριθμό κλάσεων.

**SPLITING THE DATASET**

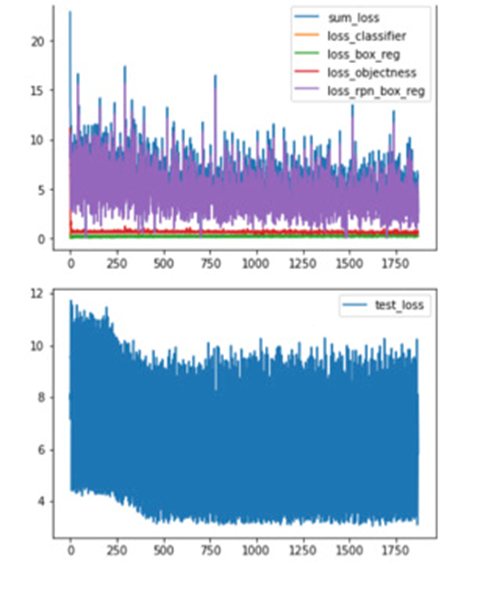
Στη πρώτη προσέγγιση επιλέχθηκε να μη γίνει καποιο augmentation. Mετέπειτα λόγω των λίγων δεδομένων έφαρμόστηκα augmentations προσεκτικά ώστε να είναι αληθοφανής οι εικόνες(καθρεπτισμοί μόνο) στα train δεδομένα ενώ ο διαχωρισμός του dataset έχει γίνει σε train και test set (80 samples) .Επιπλέον έγινε δοκιμή να γίνει augmentation μόνο σε εικόνες που περιέχουν την κλάση mask\_weard\_incorrect στο το train set .Επίσης το batch size οριστικοποιήθηκε στο 6 σε όλες τις δοκιμές αυτές .Έχει γίνει και shuffling καθώς και resize για να μπορούν να γίνουν feed στο μοντέλο.Eπίσης χρησιμοποείται η συνάρτη collate\_fn ώστε να διασφαλιστεί ότι όταν θα λαμβάνονται batches όπου όλες οι εικόνες ομαδοποιούνται μαζί και όλα τα targets ( bounding boxes και labelling) ομαδοποιούνται επίσης μαζί.

**TRAINING-EVALUATION**

Ο optimizer του μοντέλου είναι ο stochastic gradient descent, το learning rate μετά από δοκιμές είναι 0.001.Αντίστοιχα έχει επιλεχθεί το momentum=0.9 και για σε μεγαλες αλλαγές βαρών για να αποφευχθει το overfitting εχει τοποθετηθεί ένα αντιστοίχο πέναλτι μεσώ της κανονιστικής παραμετρού weight\_decay=0.0005.Oι εποχές που δοκιμάστηκαν είναι 2,3,4 και κάθε φορά με διαφορετικό iterations 100,150,200.

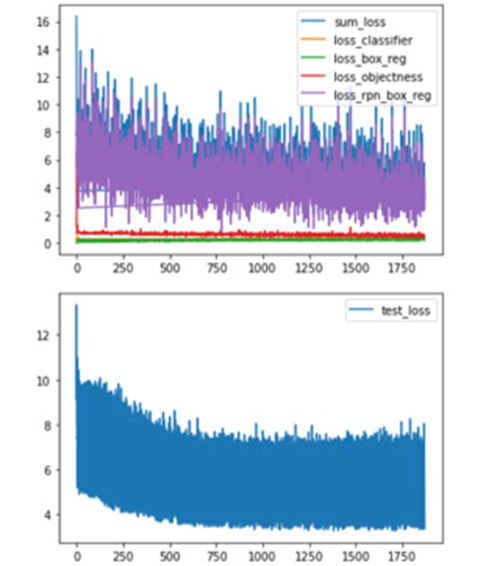
**RESULTS**

1. Χωρίς augmentations:



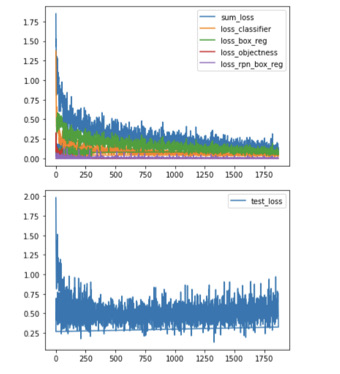
Τα losses είναι υψηλά και μάλιστα μετά από κάποιες επαναλήψεις βλέπουμε ότι σταδιακά αυξάνεται.

2)Τα αποτελέσματα για 2 εποχές και 150 iterations με augmentations στο train set είναι τα εξής:



Παρατηρείται ότι ενώ μετά από κάποια iterations το συνολικό loss πέφτει , παραμένει υψηλό.

3)Ενώ για 2 εποχές για 150 iterations αλλά με augmentation μόνο στα labels (mask\_weared\_incorrectly) που ήταν η κλάση με τα λιγότερα δεδομένα παρατηρουνταί πολύ καλύτερα αποτελέσματα.



Έτσι με threshold=0.5 φαίνονται πως τοποθετήθηκαν τα frames από το μοντέλο στις αντίστοιχες εικόνες





Όπως φαίνεται το μοντέλο τα έχει πάει πολύ καλά με εξαίρεση ένα λαθος στην κατω αριστερά φωτογραφία που αναγνώζει ότι εκεί υπαρχει πρόσωπο που δεν φοραεί σώστα τη μάσκα ενώ στη πραγματικότητα δε φοράει καθόλου μάσκα.