

Βαθιά Μάθηση και Ανάλυση Πολυμεσικών Δεδομένων

Δεύτερη Εργασία

Κωστινούδης Ευάγγελος
ΑΕΜ: 112

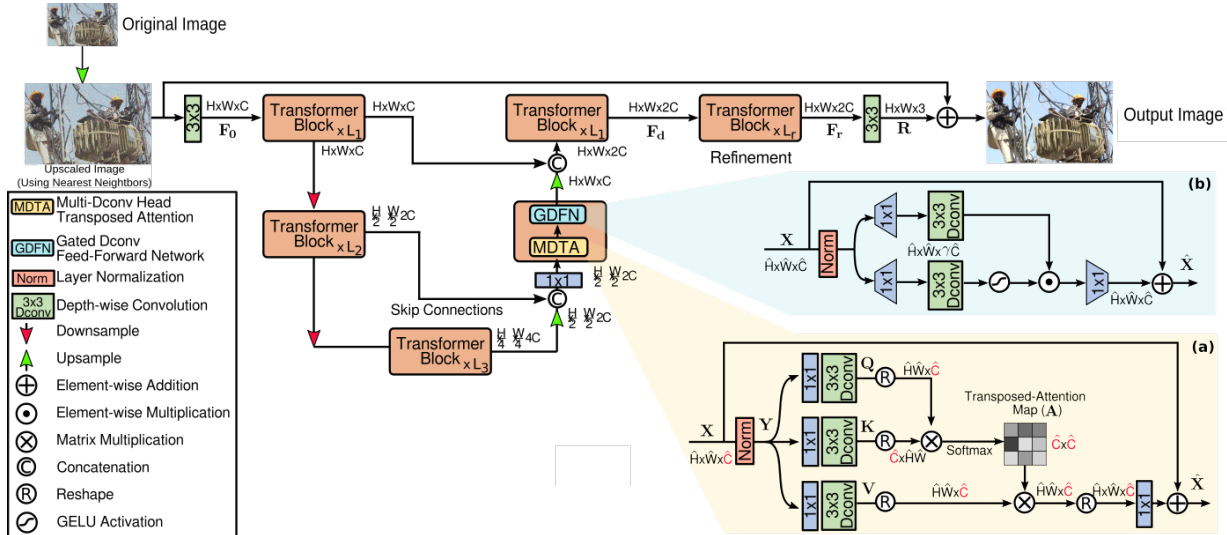
27 Μαΐου 2023

1 Περιγραφή προβλήματος που επιλέχτηκε

Για την εργασία αυτή επιλέχτηκε το πρόβλημα του super resolution. Τα δεδομένα που χρησιμοποιήθηκαν προέρχονται από τη βάση DIV2K. Συγκεκριμένα, χρησιμοποιήθηκε ένα υποσύνολο των δεδομένων αυτών.

2 Αρχιτεκτονική μοντέλου

Η αρχιτεκτονική του transformer βασίζεται στην αρχιτεκτονική του Restormer [Zam+22]. Το δίκτυο του Restormer χρησιμοποιήθηκε για την αποκατάσταση εικόνων και είναι πλήρως συνελκτικό. Η αρχιτεκτονική του δικτύου παρουσιάζεται στο παρακάτω σχήμα:



Σχήμα 1: Διάγραμμα αρχιτεκτονικής δικτύου.

Αρχικά γίνεται υπερδειγματοληψία της εικόνα εισόδου με τη μέθοδο των πλησιέστερων γειτόνων. Έπειτα, χρησιμοποιείται συνέλιξη με 3×3 kernel ούτως ώστε, από τα 3 channels της εικόνας να γίνουν C (τιμή που χρησιμοποιήθηκε είναι $C = 8$). Μετά χρησιμοποιούνται 3 επίπεδα με transformer blocks, όπου σε κάθε επίπεδο γίνεται υποδειγματοληψία μέσω pixel unshuffle (όταν το επίπεδο μεγαλώνει) και υπερδειγματοληψία μέσω pixel shuffle (όταν το επίπεδο μικραίνει). Το αποτέλεσμα που παράγουν αυτά τα επίπεδα εισέρχεται σε ένα refinement transformer block και το αποτέλεσμα του σε συνέλιξη με 3×3 kernel ούτως ώστε να υπάρχουν πάλι 3 κανάλια.

Το transformer block αποτελείται από δύο κομμάτια. Το multi Dconv head transposed attention το οποίο πραγματοποιεί τον μηχανισμό του attention μεταξύ των καναλιών (όπως φαίνεται στο Σχήμα 1). Και το gated Dconv feed-forward network.

Οι τιμές για τον αριθμό των blocks (πόσα transformer blocks υπάρχουν σε κάθε επίπεδο) και των αριθμό των heads που χρησιμοποιήθηκαν στο τελικό δίκτυο είναι:

Layer	Blocks #	Heads #
1	2	1
2	2	2
3	3	4
Refinement	2	1

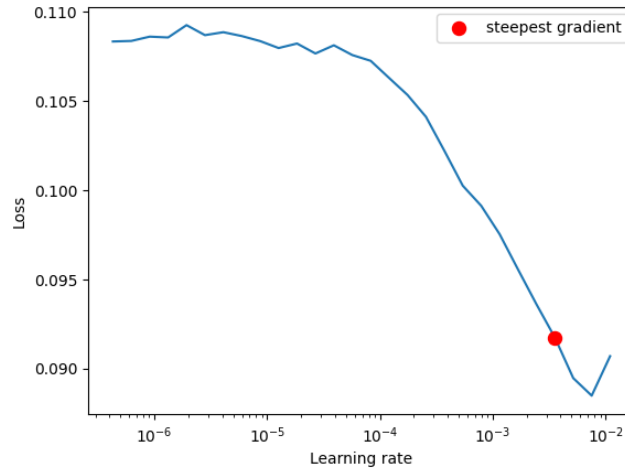
Ο αριθμός των παραμέτρων του δικτύου είναι 114.290.

3 Εκπαίδευση δικτύου

Τα δεδομένα που χρησιμοποιήθηκαν για την εκπαίδευση του δικτύου είναι οι πρώτες 100 εικόνες εκπαίδευσης της βάσης. Επίσης, για επικύρωση χρησιμοποιήθηκαν οι πρώτες 20 εικόνες επικύρωσης της βάσης.

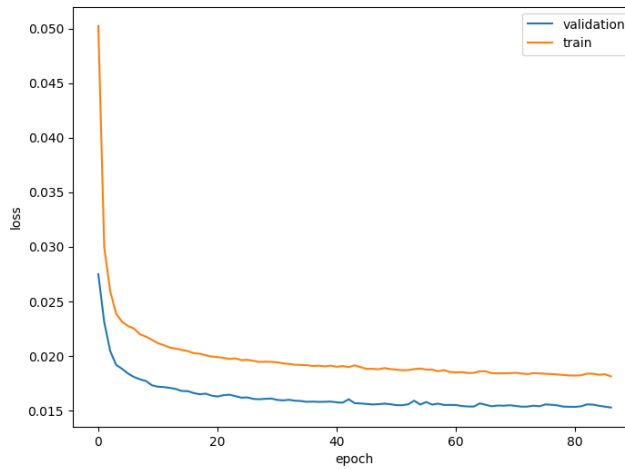
Για να έχουν όλες οι εικόνες το ίδιο μέγεθος, περικόπηκαν ούτως ώστε να έχουν μέγεθος 512×512 .

Ως συνάρτηση κόστους επιλέχτηκε το μέσο σφάλμα (L1 νόρμα). Το μέγεθος του batch που χρησιμοποιήθηκε είναι 4. Το δίκτυο εκπαιδεύτηκε για 90 εποχές. Ο αλγόριθμος βελτιστοποίησης είναι ο ADAM με ρυθμό μάθησης 0.00356. Η μέθοδος για την επιλογή του ρυθμού μάθησης βασίζεται στο [Smi17]. Σύμφωνα με αυτή τη μέθοδο, η βέλτιστη τιμή του ρυθμού μάθησης δίνεται για την τιμή με μεγαλύτερη αρνητική κλίση στο παρακάτω διάγραμμα.



Σχήμα 2: Διάγραμμα για την επιλογή του ρυθμού μάθησης.

Τα σφάλματα εκπαίδευσης του δικτύου είναι:



Σχήμα 3: Σφάλματα εκπαίδευσης και επικύρωσης κατά την διάρκεια της εκπαίδευσης.

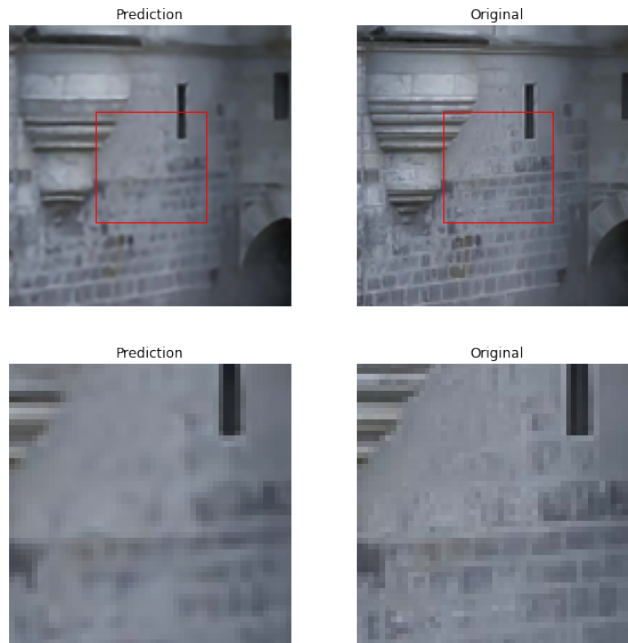
4 Αποτελέσματα

Ως σύνολο ελέγχου της απόδοσης του δικτύου χρησιμοποιήθηκαν οι τελευταίες 80 εικόνες επικύρωσης της βάσης (η βάση δεν προσφέρει δεδομένα ελέγχου). Για τις εικόνες αυτές παρουσιάζονται μετρικές για διαφορετικό μέγεθος της εικόνας εισόδου (το δίκτυο εκπαιδεύτηκε σε εικόνες 512):

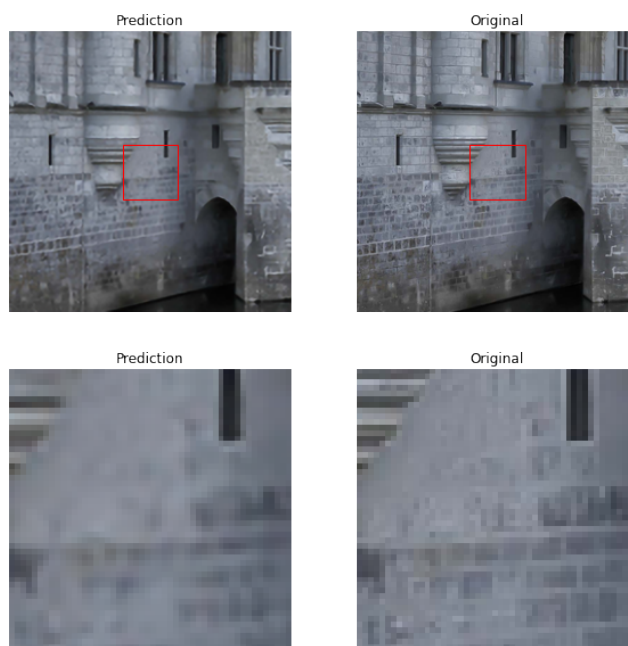
Image size	PSNR	MAE	MSE
128	31.6798	0.020571	0.001465
256	31.4181	0.019990	0.001329
512	31.2265	0.019504	0.001252
1024	31.9781	0.017348	0.001058

Παρατηρούμε ότι για είσοδο 1024 το αποτέλεσμα είναι καλύτερο για όλες τις μετρικές. Η διαφορά μεταξύ των αποτελεσμάτων δεν είναι μεγάλη.

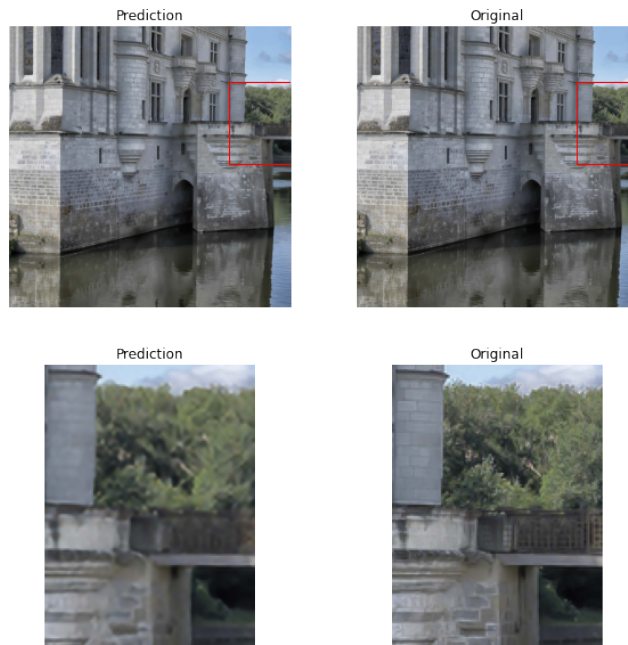
Στα [Σχήμα 4](#) [Σχήμα 5](#) [Σχήμα 6](#) φαίνονται τα αποτελέσματα του μοντέλου για διαφορετικά μεγέθη την εικόνας. Παρατηρούμε ότι η εικόνα που παράγει το μοντέλο για όλα τα μεγέθη είναι πιο θολή. Η διαφορά αυτή παρατηρείται μόνο όταν γίνει το ζουμάρισμα.



Σχήμα 4: Αποτελέσματα μοντέλου σε μία εικόνα ελέγχου μεγέθους 128×128 . Αριστερά είναι το αποτέλεσμα του μοντέλου και δεξιά η πραγματική εικόνα.



Σχήμα 5: Αποτελέσματα μοντέλου σε μία εικόνα ελέγχου μεγέθους 256×256 . Αριστερά είναι το αποτέλεσμα του μοντέλου και δεξιά η πραγματική εικόνα.



Σχήμα 6: Αποτελέσματα μοντέλου σε μία εικόνα ελέγχου μεγέθους 512×512 . Αριστερά είναι το αποτέλεσμα του μοντέλου και δεξιά η πραγματική εικόνα.

Βιβλιογραφία

- [Smi17] Leslie N Smith. “Cyclical learning rates for training neural networks.” In: *2017 IEEE winter conference on applications of computer vision (WACV)*. IEEE. 2017, pp. 464–472.
- [Zam+22] Syed Waqas Zamir et al. *Restormer: Efficient Transformer for High-Resolution Image Restoration*. 2022. arXiv: [2111.09881](https://arxiv.org/abs/2111.09881) [[cs.CV](#)].