

Βαθιά Μάθηση και Ανάλυση Πολυμεσικών Δεδομένων

Πρώτη Εργασία

Κωστινούδης Ευάγγελος
ΑΕΜ: 112

17 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 συνελικτικού δικτύου είναι:

1. $Conv2d(3, 128, kernel_size = (3, 3), stride = (1, 1), padding = (1, 1))$
2. $LeakyReLU(negative_slope = 0.01)$
3. $ConvTranspose2d(128, 64, kernel_size = (4, 4), stride = (2, 2), padding = (1, 1))$
4. $LeakyReLU(negative_slope = 0.01)$
5. $Conv2d(64, 32, kernel_size = (3, 3), stride = (1, 1), padding = (1, 1))$
6. $LeakyReLU(negative_slope = 0.01)$
7. $Conv2d(32, 3, kernel_size = (3, 3), stride = (1, 1), padding = (1, 1))$

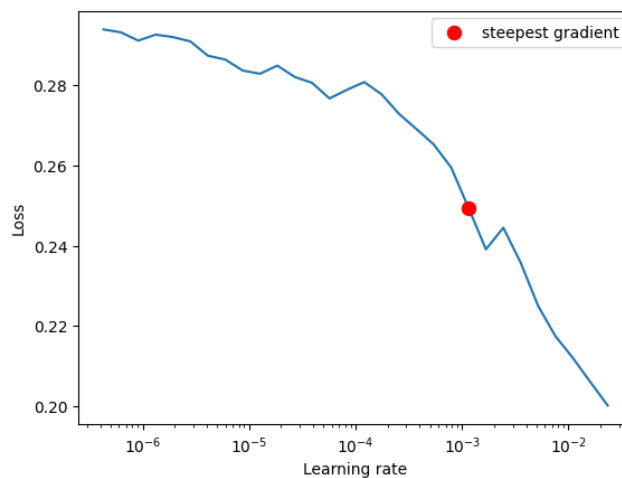
Το δίκτυο δέχεται ως είσοδο μία RGB εικόνα και παράγει στην έξοδο την ίδια εικόνα με τις διπλάσιες διαστάσεις. Αποτελείται μόνο από συνελκτικά επίπεδα, γεγονός που σημαίνει ότι το μέγεθος της εικόνας της εισόδου δεν είναι προκαθορισμένο, αλλά μπορεί να μεταβάλλεται. Χρησιμοποιείται η LeakyReLU ως συνάρτηση ενεργοποίησης.

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.

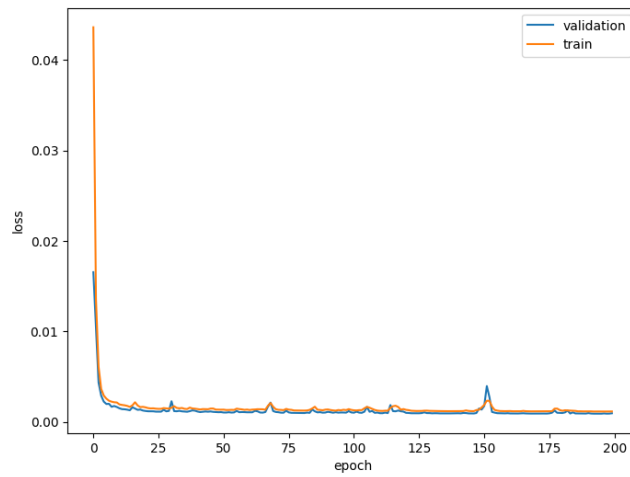
Για να έχουν όλες οι εικόνες το ίδιο μέγεθος, περικόπηκαν ούτως ώστε να έχουν μέγεθος 1000×1000 .

Ως συνάρτηση κόστους επιλέχτηκε το μέσο τετραγωνικό σφάλμα. Το μέγεθος του batch που χρησιμοποιήθηκε είναι 4. Το δίκτυο εκπαιδεύτηκε για 200 εποχές. Ο αλγόριθμος βελτιστοποίησης είναι ο ADAM με ρυθμό μάθησης 0.00115. 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 1: Διάγραμμα για την επιλογή του ρυθμού μάθησης.

Τα σφάλματα εκπαίδευσης του δικτύου είναι:

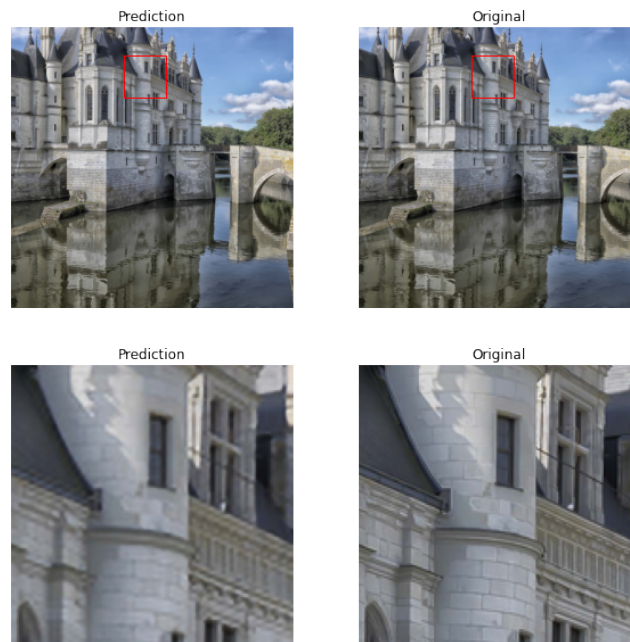


Σχήμα 2: Σφάλματα εκπαίδευσης και επικύρωσης κατά την διάρκεια της εκπαίδευσης.

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 το μέσο τετραγωνικό σφάλμα είναι 0.001231 και το μέσο PSNR είναι 31.1708.

Στο [Σχήμα 3](#) φαίνονται τα αποτελέσματα του μοντέλου. Παρατηρούμε ότι η εικόνα που παράγει στο μοντέλο είναι πιο θολή. Η διαφορά αυτή παρατηρείται μόνο όταν γίνει το ζουμάρισμα.



Σχήμα 3: Αποτελέσματα μοντέλου σε μία εικόνα ελέγχου. 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.