

ОГЛЯД SUPER-RESOLUTION CONVOLUTIONAL NEURAL NETWORK, З МЕТОЮ ВИКОРИСТАННЯ ДЛЯ ПІДВИЩЕННЯ РОЗДІЛЬНОЇ ЗДАТНОСТІ ЗОБРАЖЕНЬ

Вінницький національний технічний університет

Анотація. У статті розглянуто super-resolution convolutional neural network, з метою використання для підвищення роздільної здатності зображень.

Ключові слова: super-resolution convolutional neural network, зображення, роздільна здатність, нейронні мережі.

Вступ

З інтенсивним розвитком інформаційних і комп'ютерних технологій швидкими темпами вдосконалюється фото і відеоапаратура та розширюються області їх застосування.

На сьогоднішній день фото та відеодані використовуються в промисловості, медицині, військовій та космічній сферах, охоронних системах, мистецтві та багатьох інших галузях. Формування чітких зображень, поліпшення їх якості та автоматизація різноманітних видів обробки зображень, включаючи зображення, що створюються електронними мікроскопами, рентгенівськими апаратами, томографами, штучними супутниками, безпілотними літальними апаратами, камерами відеоспостереження тощо, є предметом сучасних досліджень та розробок. При цьому гострою залишається проблема покращення роздільної здатності.

В зв'язку з цим актуальними задачами є вдосконалення та створення нових методів цифрової обробки зображень. В даній доповіді розглянемо метод Super-Resolution Convolutional Neural Network для підвищення роздільності зображень.

Огляд методу

Даний метод підвищення роздільності зображень використовує штучні нейронні мережі, фільтри та навчання для обробки зображень.

Перед тим, як здійснювати збільшення роздільності зображення за допомогою даної моделі, зображення попередньо збільшується до необхідних розмірів. Як правило, при цьому використовується алгоритм бікубічної інтерполяції. Хоча формально даний метод інтерполяції і є згортою, а самі автори дослідження [3] стверджують, що до нього може бути застосована формулювання «шар», в описі архітектури нейронної мережі операція згортка звичайно не фігурує. Архітектура мережі SRCNN відображено на рисунку 1.

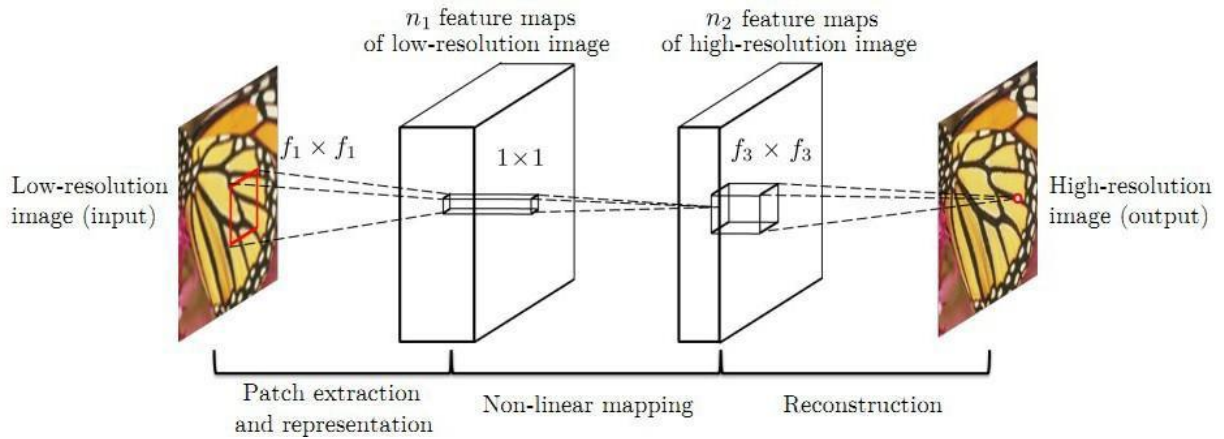


Рисунок 1 – Архітектура мережі SRCNN [3]

Позначимо вихідне зображення як Y , а зображення з високою роздільністю, з яким буде здійснюватися порівняння – X . Зображення Y , незважаючи на відповідність розширенню зображення X , будемо вважати зображенням низької роздільності. На рисунку 1 зображення Y позначено написом «Low resolution image (input)».

В результаті роботи нейронної мережі необхідно отримати зображення $F(Y)$ об'єктивно більш подібне, ніж спотворене, до вхідного зображення X . Для цього необхідно знайти функцію F . Пошук складається з наступних 3-х операцій:

1. Отримання патчів. На рисунку 1 дана операція позначена написом «Patch extraction and representation». Результатом роботи даної операції є патчі (окремі ділянки зображення, що складаються з декількох пікселів) з вхідного зображення низького розширення Y , а потім представити кожен патч у вигляді багатовимірного вектора. Ці вектори містять набір карт ознак, розмір яких дорівнює розмірності вектора (n_1 feature maps of low-resolution image).
2. Нелінійне відображення. На рисунку 1 дана операція позначена написом «Non-linear mapping». Результатом операції є нелінійне відображення кожного багатовимірного вектора на інші вектори. Кожен відображений вектор являє патч з високою роздільною здатністю. Ці вектори містять інший набір карт ознак (n_2 feature maps of high-resolution image).
3. Реконструкція. На рисунку 1 позначена написом «Reconstruction». Дана операція використовує патча, отримані з нелінійного відображення використовуються для того, щоб згенерувати зображення високої роздільної здатності $F(Y)$. Очікується, що отримане зображення буде об'єктивно схожим з оригінальним зображенням X .

Отримане в результаті роботи нейронної мережі зображення $F(Y)$ позначено на рисунку 1 написом «High-resolution image (output)». Всього дана модель містить три шари.

Для отримання патчів використовується перший шар нейронної мережі, який можна позначити (1):

$$F_1(Y) = \max(0, W_1 * Y + B_1), \quad (1)$$

де W_1 – фільтри,
 B_1 – ваги,
 $*$ – операція згортки.

Фільтри W_1 можна позначити через (2):

$$W_1 = c \times f_1 \times f_1, \quad (2)$$

де c – кількість каналів зображення Y ,
 f_1 – розмір фільтра у просторі.

Фільтри W_1 здійснюють n_1 операцій згортки зображення, і кожна згортка має ядро згортки розміру $c \times f_1 \times f_1$. Вихідні дані шару містять карти ознак в кількості n_1 . Ваги B_1 це n_1 -мірний вектор, кожен елемент якого співставлено з елементом фільтра W_1 .

Як функції активації першого шару використовується ReLU (Rectified Linear Unit, «випрямляч») (3):

$$f(x) = \max(0, x), \quad (3)$$

де x – вхідний сигнал.

Другий шар здійснює операцію нелінійного відображення n_1 -мірний векторів на n_2 -мірні вектори. Операція працює лише з фільтрами розміром 1×1 . Операція, яка здійснюється другим шаром ШНМ, математично описується наступним чином в (4):

$$F_2(Y) = \max(0, W_2 * F_1(Y) + B_2), \quad (4)$$

де параметри W_2 і B_2 аналогічні параметрам для F_1 , але фільтри тепер здійснюють n_2 операцій, а вагами є n_2 -мірний вектор.

Подібно до першого шару в даному випадку в якості функції активації також використовується ReLU. Третій шар здійснює операцію реконструкції. Його можна представити (5):

$$F(Y) = W_3 * F_2(Y) + B_3, \quad (5)$$

де W_3 – це фільтри розмірністю $n_2 \times f_3 \times f_3$,
 B_3 – c -мірний вектор.

На виході після обробки даних третім шаром виходить зображення з високою роздільною здатністю.

Функція активації даного шару (6):

$$f(x) = x, \quad (6)$$

де x – це вхідний сигнал.

В оцінці втрат використовується середня квадратична помилка MSE (Mean Squared Error) (7):

$$L(\theta) = \sum_{i=1}^n |F * (Y_i) - X_i|, \quad (7)$$

де θ – це параметри нейронної мережі $\{W_1, W_2, W_3, B_1, B_2, B_3\}$,
 n – кількість зразків в навчальній вибірці.

При аналізованні даної моделі було проведено тестування для різних значень вхідних даних для нейронної мережі. Було з'ясовано, що найбільш оптимальними є наступні вхідні дані: $f_1 = 9, f_2 = 1, f_3 = 5, n_1 = 64$ і $n_2 = 32$ [3].

Висновки

В даній доповіді було розглянуто super-resolution convolutional neural network як спосіб збільшення роздільної здатності зображень. Детально оглянувши теоретичне підґрунтя для роботи цього методу, впевнилися, що його можна використовувати для збільшення роздільної здатності зображень.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. R. Keys. Cubic convolution interpolation for digital image processing [Text] // IEEE Transactions on Acoustics, Speech, and Signal Processing. 29 (6): 1153—1160.
2. Hwang J. W., Lee H. S. Adaptive image interpolation based on local gradient features [Text] // IEEE Signal Processing Letters. – 2004. – Т. 11. – 1. 3. – С. 359-362.
3. Dong C. et al. Image super-resolution using deep convolutional networks [Text] // IEEE transactions on pattern analysis and machine intelligence. – 2016. – Т. 38. – 1. 2. – С. 295-307.
4. Romano Y., Isidoro J., Milanfar P. RAISR: rapid and accurate image super resolution [Text] // IEEE Transactions on Computational Imaging. – 2017. – Т. 3. – 1. 1. – С. 110-125.
5. The Most Popular Language For Machine Learning and Data Science Is [Online]. Available : <http://www.kdnuggets.com/2017/01/most-popular-language-machine-learning-data-science.html>
6. Бібліотека глибокого навчання Tensorflow. [Online] Available :

<https://habrahabr.ru/company/ods/blog/324898/>

Білоус Антон Олександрович – студент групи ІАКІТ-186, факультет автоматизації на комп'ютерно інтегрованих технологій, Вінницький національний технічний університет, м. Вінниця.

Софіна Ольга Юрївна — кандидат технічних наук, доцент кафедри автоматизації та інтелектуальних інформаційних технологій, Вінницький національний технічний університет, м. Вінниця.