

ПІДХІД ДО РОЗРОБКИ НЕЙРОННОЇ МЕРЕЖИ ТРАНСФОРМАЦІЇ ЗОБРАЖЕНЬ

Левченко Вячеслав Ігорович

Тішин Петро Металлинович, к.т.н., доцент

Багато класичних проблем сучасності можуть бути обрешені як завдання перетворення зображення, де система отримує деякий вхідний образ і перетворює його у вихідний образ. Приклади обробки зображень включають розпізнавання, надрозміри та кольорову обробку, де вхід - це деградоване зображення (шум, низька роздільна здатність або відтінки сірого) та вихід - високоякісний кольоровий образ. Приклади комп'ютерного бачення включають семантичну сегментацію та оцінку глибини, де вхідні дані є кольоровим зображенням і вихідне зображення кодує семантичну або геометричну інформацію про сцену.

Одним з підходів до вирішення завдань трансформації та обробки зображень є навчання підконтрольної нейронної мережі з належним напрямком під наглядом, використовуючи для кожного пікселя функцію втрати для вимірювання різниці між вихідними і існуючими зображеннями. Останні вивчені методи для вирішення таких проблем зазвичай використовують згорткові нейронні мережі зі зворотним зв'язком, використовуючи втрату в пікселях між вихідними і вхідними зображеннями. Такі підходи є ефективні під час випробувань, вимагаючи лише пропуск через навчену мережу. Проте, втрати за піксель, які використовуються цими методами, не фіксують перцептуальні відмінності між вихідними та істинними зображеннями. Наприклад, розглядавши два однакових зображення, зміщені один за одним на один піксель, незважаючи на їх сприйняття подібності для людини, вони будуть дуже різними, якщо виміряти різницю за втратами в пікселях.

Для вирішення проблеми швидкості обробки запропоновано проводити підготовку трансформаційних мереж для обробки зображень, але замість використання функцій втрати в пікселях залежно від інформації про пікселі низького рівня, навчаємо мережі за допомогою функцій перцептивної збитковості, які залежать від функцій високого рівня від мережі попередньо запобіжних втрат. Під час тренування перцептивні втрати подібності зображення набагато сильніше, ніж втрати на піксель, і на час тестування трансформаційні мережі працюють у режимі реального часу. Як показано на рисунку 1, система обробки зображень складається з двох компонентів: мережі перетворення зображень f_w та мережі витрат ϕ , що використовується для визначення деяких функцій витрат l_1, \dots, l_k . Мережа трансформації зображення - це глибока залишкова нейронна сверточна мережа, параметризована вагами W ; вона перетворює вхідні зображення x у вихідні зображення y' за допомогою функції $y' = f_w(x)$. Кожна функція втрат обчислює скалярне значення $l_k(y', y_k)$, яке вимірює різницю між вихідним зображенням y' з цільовим зображенням y_k . Мережа перетворення зображень навчається, використовуючи стохастичний градієнтний спуск, щоб мінімізувати зважену комбінацію функцій втрат.

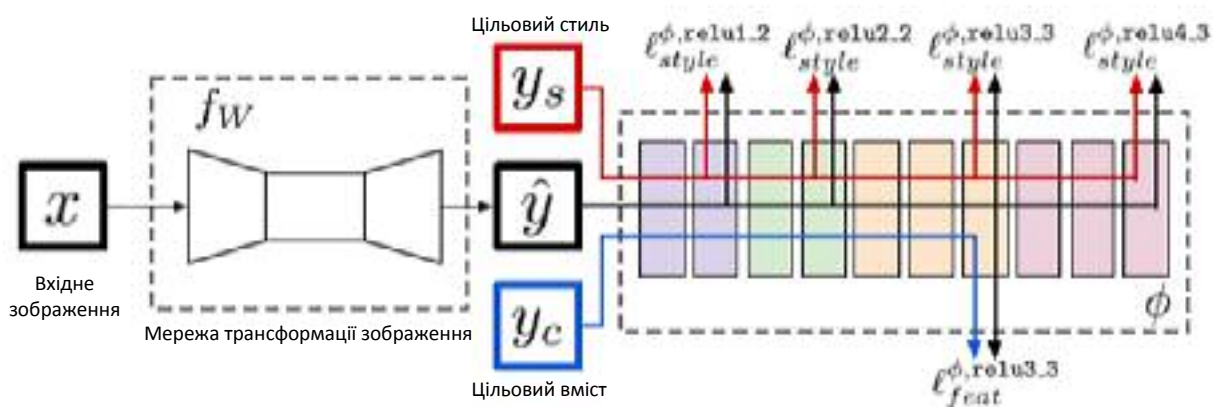


Рис. 1. Огляд системи

Щоб усунути недоліки втрат у пікселях і дозволити функціям втрати краще виміряти перцептивні та семантичні відмінності між зображеннями, був використаний принцип, який генерує зображення за допомогою оптимізації. Ключовою ознакою цих методів є те, що сформовані нейронні мережі, попередньо підготовлені для класифікації зображень, вже навчилися кодувати перцептуальну та семантичну інформацію, яку треба виміряти в функціях втрати. Тому використовували мережу Φ , яка попередньо підготовлена для класифікації зображень як мережа з фіксованими втратами, з метою зниження функцій втрат. Глибока конволюційна трансформаційна мережа навчається за допомогою функцій втрат.

Мережа втрат використовується для визначення втрат реконструкції об'єкта I_{feat}^{Φ} та втрат відновлення стилю I_{style}^{Φ} , які вимірюють відмінності у вмісті та стилі між зображеннями. Для кожного вхідного зображення x ми маємо цільовий вміст y_c і ціль стилю y_s . Для передачі стилю ціль контенту y_c є вхідним зображенням, а вихідне зображення y' має поєднувати вміст $x = y_c$ зі стилем y_s ; також тринуємо одну мережу за стильовою ціллю. Для надвисокої роздільної здатності з одним зображенням вхідне зображення x - це вхід із низькою роздільною здатністю, цільовий вміст y_c є зображенням з високою роздільною здатністю на основі істини, а втрата відновлення стилю не використовується; ми тренуємо одну мережу за коефіцієнтом супер-дозволу.

Описаний підхід можна використовувати для навчання та донавчання мережі за допомогою використання функцій втрати сприйняття для інших завдань трансформації зображень, таких як кольорова та семантична сегментація. Основна увага буде спрямована на навчання мережі за допомогою використання мережі різноманітних втрат, щоб побачити, наприклад, що мережі втрат, навчені різним завданням або набором даних, можуть передавати результати до мережі перетворення зображень з різними типами семантичних знань.