

ПОШУК ОБ'ЄКТІВ У ВІДЕОПОТОЦІ ЗА ДОПОМОГОЮ ЗГОРТКОВИХ НЕЙРОННИХ МЕРЕЖ

Запорізька державна інженерна академія, кафедра ПЗАС

Останнім часом у штучному інтелекті набирають популярності напрямки, які базуються на біологічних принципах. Навіть можна спрогнозувати, що незабаром більш просунуті моделі штучного інтелекту набудуть широкого використання. Одна з основних рушійних сил в цьому напрямку - це нові алгоритми навчання, що дозволяють обробляти великі обсяги інформації. Сучасні розробки в області алгоритмів машинного навчання та комп'ютерного зору можуть швидко приймати рішення з точністю не гірше ніж професіонал.

Використання класичних нейронних мереж для розпізнавання зображень ускладнено великою розмірністю вектора вхідних значень нейронної мережі, великою кількістю нейронів в проміжних шарах і, як наслідок, великими витратами обчислювальних ресурсів на навчання і обчислення мережі. Згортковим нейронним мережам в меншій мірі властиві такі недоліки.

Згорткова нейронна мережа (англ. convolutional neural network, CNN) — спеціальна архітектура штучних нейронних мереж, входить до технології глибокого навчання (англ. deep learning) [5]. Ця технологія побудована за аналогією з принципами роботи зорової кори головного мозку, в якій були відкриті так звані прості клітини, що реагують на прямі лінії під різними кутами, і складні клітини, реакція яких пов'язана з активацією певного набору простих клітин [5]. Ці мережі було натхнено біологічними процесами, вони є варіаціями багатошарових перцептронів, розробленими для використання мінімальної попередньої обробки. Згорткові мережі мають широке застосування в розпізнаванні зображень та відео, рекомендаційних системах та обробці природної мови.

На відміну від класичних нейронних мереж, на вхід окремого шару потрібно подавати тривимірну матрицю, проте на виході із шару її розмірність може бути змінена. Найбільш популярними є наступні типи шарів: вхідний, згортковий, шари субдискретизації, повноз'єднаний [3].

Зазвичай *вхідний шар* лише трансформує двовимірне зображення у тривимірну матрицю. Зріз отриманої матриці має такі самі розміри як і зображення, проте містить інформацію лише про один із трьох компонентів кольору, отже глибина вихідної матриці буде дорівнювати 3.

Згортковий шар — це основний компонент цього типу штучних нейронних мереж. Цей шар здійснює згортку результату попереднього шару. Кожен такий шар повинен мати наступні характеристики:

1. *розмір рецептивного поля F* відповідає за кількість нейронів попереднього шару, які будуть зв'язані з одним нейроном у цьому шарі;
 - *глибина K* відповідає за кількість нейронів у цьому шарі (скільки нейронів, такою і буде глибина матриці на виході);
 - *крок S* контролює як сильно будуть перекриватися рецептивні поля;
 - *розмір нульового доповнення P* дозволяє контролювати розмір вихідної матриці — залежно від нього вхід на краях буде доповнено нулями.

Кожен нейрон у цьому шарі представлений фільтром (ядром) — матрицею розміром $F \times F \times F$, її параметри змінюються під час навчання. За допомогою цього фільтра формується двовимірний карт активзації, яка є зрізом вихідної матриці шару.

Шар субдискретизації змінює розмірність карти ознак, це дозволяє спростити обчислення а також може запобігати перенавчанню. Метод максимізаційної підвиборки показав найкращі практичні результати. Його ідея полягає у тому що ми для кожного зрізу

вхідної матриці застосовуємо фільтр розміром 2×2 : заміняємо 4 числа на максимальне з них, та на виході отримуємо матрицю з меншою у 2 рази розмірністю [2].

Також важливим є *повнозв'язний шар* за допомогою якого визначається клас на основі отриманих ознак. Цей шар працює повністю аналогічно шару з класичних штучних нейронних мереж.

Як показала практика та експерименти, згорткові нейронні мережі досить точно виконують розпізнавання образів на статичних зображеннях. Тому доречно працювати з відео потоком як з набором статичних кадрів. У якості оптимізації було використано емпіричне правило: пошук об'єкта виконувався на краях кадру, а також поруч з тим місцем де його було знайдено на попередньому кадрі [1,4].

У якості проблеми було обрано пошук та визначення гравців на відео. Для цього було використано каскад згорткових нейронних мереж: мережа для пошуку гравців та мережа для визначення який саме це гравець. Нейронна мережа для пошуку гравців повинна мати незначну похибку, а також вона повинна відрізнити саме гравців від інших людей, наприклад судді. Нейронна мережа для визначення гравця, на відміну від першої, має право на неоднозначну відповідь. Це викликано наступними факторами:

2. іноді навіть людина не може визначати гравця;
 - якщо мережа не може дати однозначну відповідь, їй може допомогти людина;
 - навіть якщо у даному кадрі неможливо однозначно визначити гравця, існує ймовірність, що це можна зробити за допомогою інших кадрів, і це вже робота алгоритмічної системи яка стежить за пересуванням гравців.

Висновки

За час дослідження проблеми було вивчено перспективний напрямок розвитку штучних нейронних мереж, а саме згорткові нейронні мережі.

Література

1. M.L. Eichner, T.P. Breckon, "Integrated Speed Limit Detection and Recognition from Real-Time Video", Proc. IEEE Intelligent Vehicle Symposium, The Netherlands, 2008
1. Patrice Y. Simard, Dave Steinkraus, John C. Platt Best Practices for Convolutional Neural Networks Applied to Visual Document Analysis
2. A Tutorial on Deep Learning Part 2: Autoencoders, Convolutional Neural Networks and Recurrent Neural Networks - 2015.
3. N. Lagunov The Use of Convolutional Neural Networks for Multiparameter Objects Recognition - 2013.
4. Інтернет ресурс: <http://cs231n.github.io/convolutional-networks/>