

Нарощування даних в задачі детекції об'єктів

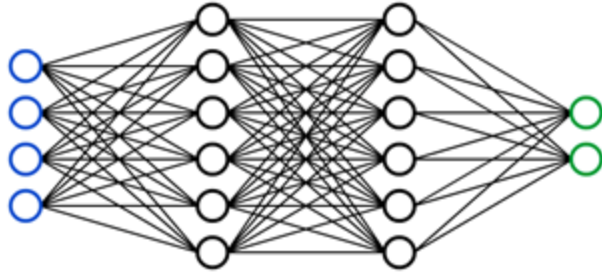


NATIONAL UNIVERSITY OF
KYIV-MOHYLA ACADEMY

Решетар Олександр
Прикладна математика 4 курс
Керівник: Швай Надія Олександрівна

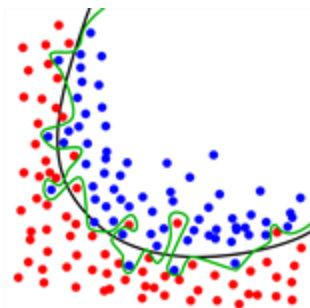
Актуальність проблеми

Для тренування нейронної мережі необхідна велика кількість даних: чим більше тим краще. Яскравим прикладом є продукт від компанії OpenAI - Chat GPT. Загальний об'єм тренувальних даних сягнув 45TB!



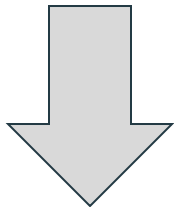
Актуальність проблеми

Щоб забезпечити стійкість мережі до викидів, запобігти перевивченню та мати доступ до розмаїття даних, використовують методи аугментації. Аугментація даних — техніку штучного збільшення навчального набору шляхом створення модифікованих копій набору даних з використанням існуючих даних. Аугментація може застосовуватися до будь-якого типу даних – від чисел до зображень. Вона є важливим інструментом у сфері нейронних мереж. Вона має кілька переваг, таких як регуляризація моделі та збільшення розміру набору даних. Особливо це важливо у випадку датасетів, де отримання міток для даних є дорогим процесом.



Види аугментацій

Прості



Стиск, розтяг,
повороти, змінення
кольорової гами

Translate

$$\begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}$$



Scale

$$\begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$



Rotate

$$\begin{bmatrix} c & s & 0 \\ -s & c & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$c = s = \sin(45^\circ)$$

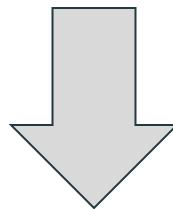


Shear

$$\begin{bmatrix} 1 & 0.5 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}$$



Складні



Мозаїка, Пуассонівська
вставка, генерація нових
семплів

Дослідження

В даній роботі було розглянуті складні методи аугментації, які дозволяють суттєво збільшити розмір датасету. В той же час, їхня реалізація доволі проста, що дає можливість генерувати нові вибірки швидко. До них належать:

- copy-paste
- segmented copy-paste
- gaussian blur
- poisson image blending
- cp-gan

Для тренування було обрано датасет Pascal VOC через його різноманітність анотованих даних. В якості моделі було застосовано YOLOv5.



Copy-paste vs segmented copy-paste

Найпростіший спосіб - звичайна вставка об'єкта у цільове зображення з маскою

$$I(x, y) = M(x, y)S(x, y) + (1 - M(x, y))T(x, y) = \begin{cases} S(x, y), & M(x, y) = 1 \\ T(x, y), & M(x, y) = 0 \end{cases}$$

де M - маска, S - вихідне зображення, T - цільове зображення



Проста вставка



Сегментована вставка

Гауссовий блюр

Вдосконалений варіант для плавного переходу між контурами вставленого об'єкту та цільовим зображенням.

$$I(x, y) = f(M)S(x, y) + (1 - f(M))T(x, y)$$

де $f(M)$ - коефіцієнт блюру, M - маска, S - вихідне зображення, T - цільове зображення.

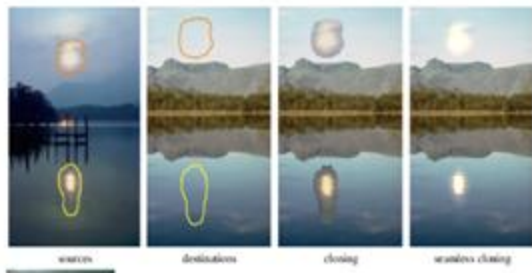


Пуассонівська вставка

Використовуючи рівняння пуассона, можна згладити контури вставленого об'єкту, досягаючи таким чином плавний перехід.

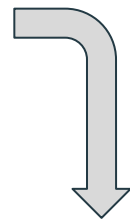
$$\nabla^2 I(x, y) = \nabla^2 S(x, y) \in \Omega$$

$$s.t. I(x, y) = T(x, y) \in \partial\Omega$$



0	1	0
1	-4	1
0	1	0

Оператор Лапласа



Даний фільтр розпізнає края об'єктів. Проходячи ним по вставленому зображенню, можна зробити плавний перехід між кольоровою гамою.

$$\begin{aligned} & -4I(x, y) + I(x + 1, y) + I(x, y + 1) + I(x - 1, y) + I(x, y - 1) \\ & = -4S(x, y) + S(x + 1, y) + S(x, y + 1) + S(x - 1, y) + S(x, y - 1) \end{aligned}$$

Пуассонівська вставка

Є два випадки розташування пікселів: на межі та всередині площини.

$$\begin{aligned} & -4I(x, y) + I(x + 1, y) + I(x, y + 1) + I(x - 1, y) + I(x, y - 1) \\ & = -4S(x, y) + S(x + 1, y) + S(x, y + 1) + S(x - 1, y) + S(x, y - 1) \end{aligned}$$



В цьому випадку в нас лінійне рівняння з 5-ма невідомими



Оскільки границі нам відомі(цільове зображення), рівняння сильно спрощується



$$\begin{aligned} I(x-1, y) &= T(x-1, y) \\ I(x, y-1) &= T(x, y-1) \end{aligned}$$

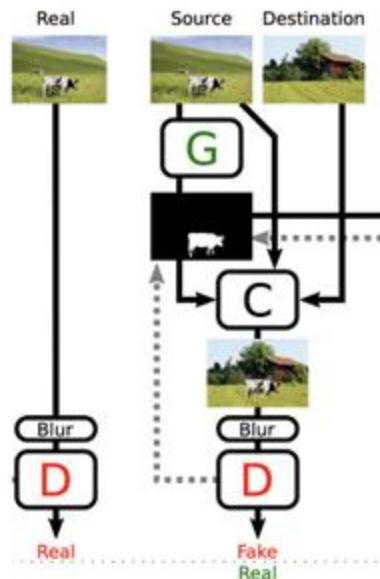
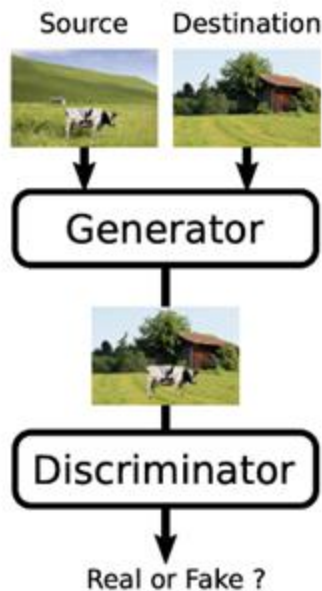
GAN -> CP-GAN

Генеративна змагальна мережа - новий підхід у сфері нейронних мереж, де за один підхід навчаються дві мережі, які змагаються одна з одною

$$\min_G \max_D V(D, G)$$

$$V(D, G) = \mathbb{E}_{x \sim p_{data}(x)} [\log D(x)] + \mathbb{E}_{z \sim p_z(z)} [\log(1 - D(G(z)))]$$

Генератор мінімізує $\log(1 - D(G(z)))$, тобто генерує таке зображення, щоб дискримінатор визначив його як істинне. Дискримінатор навпаки - максимізує даний вираз.



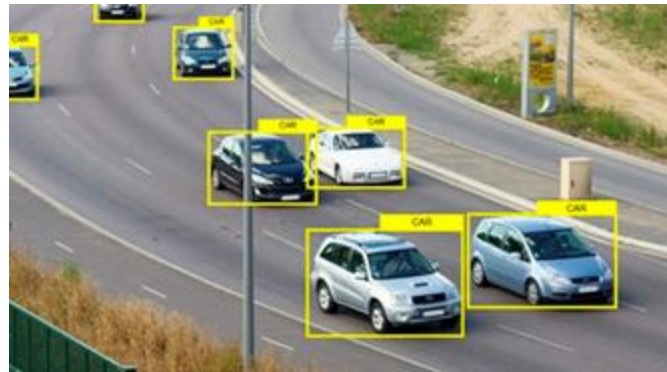
Алгоритм вставки

Algorithm 1 Алгоритм вставки

Require: $(x_0, y_0) \geq 0, (x_1, y_1) \leq \text{img.width}, \text{img.height}$

```
scale ← 100
overlap ← 0.05
points ← []
while points.size = 0 do
  if scale < 50 then
    overlap ← overlap + 0.05
    scale ← 100
  end if
  maskx ← bbox[2] - bbox[0]
  masky ← bbox[3] - bbox[1]
  for i = 0, i < width do
    for j = 0, j < height do
      x0, y0 ← i, j
      x1, y1 ← min(i + maskx, width), min(j + masky, height)

      for targetBbox ∈ targetBoxes do
        xtarget0, ytarget0 ← targetBbox[0], targetBbox[1]
        xtarget1, ytarget1 ← targetBbox[2], targetBbox[3]
        check 4 corners of the source object
      end for
      if overlap ≤ area of intersections with all target boxes then
        points.append((i, j))
      end if
    end for
  end for
end while
```



Результати

Метод	Клас	Епох	map.5	map.95
Baseline	-	50	0.83	0.589
Naive Copy-paste	dining table, bus	50	0.812	0.564
Segmented copy-paste paste	dining table, bus	50	0.81	0.561
Gaussian blur	dining table, bus	50	0.828	0.582
Poisson image editing	dining table, bus	50	0.828	0.582
Copy-GAN	dining table, bus	50	0.832	0.59

Висновки

- Найкращу точність передбачень дав метод, застосовуючи ГАНи.
- Інші методи показали таку ж точність як бейзлайн, або нижчу. Можливо варто вдосконалити алгоритм вставки об'єктів в цільове зображення таким чином, щоб градієнти контурів збігались.

Література

1. Golnaz Ghiasi, Yin Cui, Aravind Srinivas, Rui Qian, Tsung-Yi Lin, et al. Simple Copy-Paste is a Strong Data Augmentation Method for Instance Segmentation. arXiv:2012.07177, 2020
1. Alexey Bochkovskiy, Chien-Yao Wang, Hong-Yuan Mark Liao. YOLOv4: Optimal Speed and Accuracy of Object Detection, arXiv:2004.10934, 2020
2. Hongyi Zhang, Moustapha Cisse, Yann N. Dauphin, David Lopez-Paz. mixup: Beyond Empirical Risk Minimization, arXiv:1710.09412. 2018
3. Zhang, H., Cisse M., Dauphin Y. N., Lopez-Paz D. Mixup: Beyond Empirical Risk Minimization. arXiv:1710.09412, 2017
4. Lingzhi Zhang, Tarmily Wen, Jianbo Shi. Deep Image Blending, arXiv:1910.11495, 2019