徐宏民老師 - CFOCNet

出自

WACV

https://openaccess.thecvf.com/content/WACV2021/papers/Yang_Class-Agnostic_Few-Shot_Object_Counting_WACV_2021_paper.pdf

題目

CFOCNet

Class-agnostic Few-shot Object Counting

Introduction

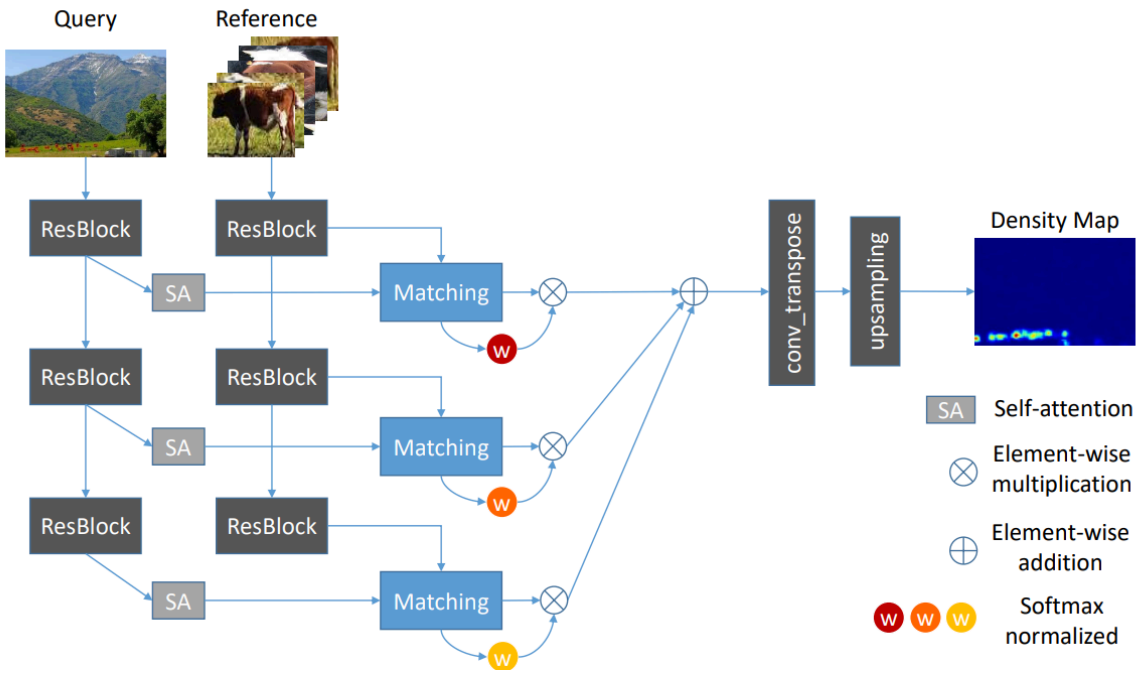
背景

medical and bioscience 領域沒辦法每個類別都做標記，很貴

動機

目的

1. 用 Resnet 抽取 class-independent 的 feature
2. reduces the counting problem to a matching problem : 以 refference 為 filter 去 filter query



相關研究

1. few-shot learning :
   training set 中有 C 個種類，每個種類有 K 張照片， given input 要分類成 C 類，此任務稱為 C-way K-shot learning

方法

訓練時只能有 K 個種類，而 TESTING 時的 T 個種類都需要與 K 個種類互相不同

所以本實驗使用 COCO 的子集

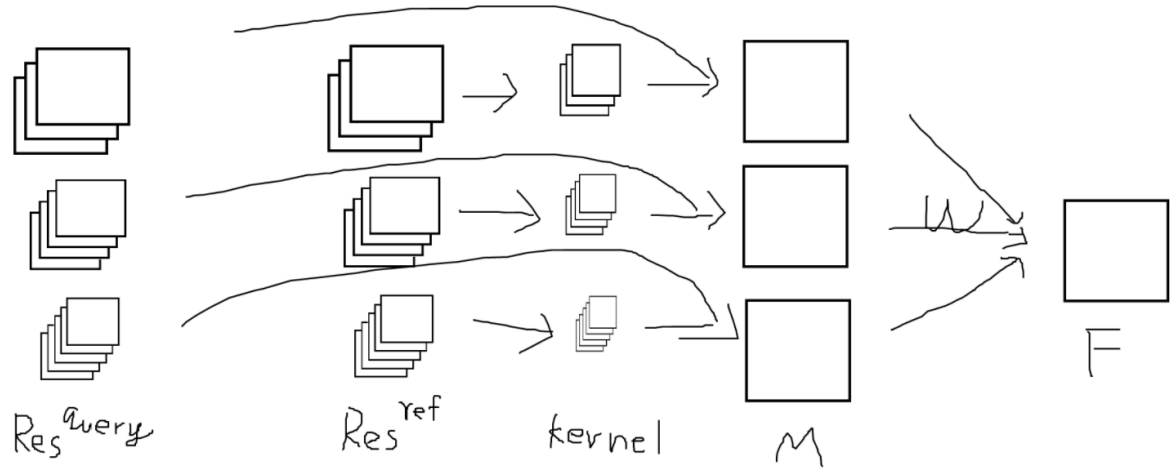
query image 中可以有任一種，而 reference image 只能包含一個種類

為了讓 refference 可以不定張，所以使用 flexible network architecture

模型架構

1. fully conv network ⇒ input size 要任意都可以
2. number of reference images needs not to be fixed
3. encoder of Request and Query 後，會做 correlation operation
4. calculates the matching score in different scales of Resnet
5. decoder generates density map.

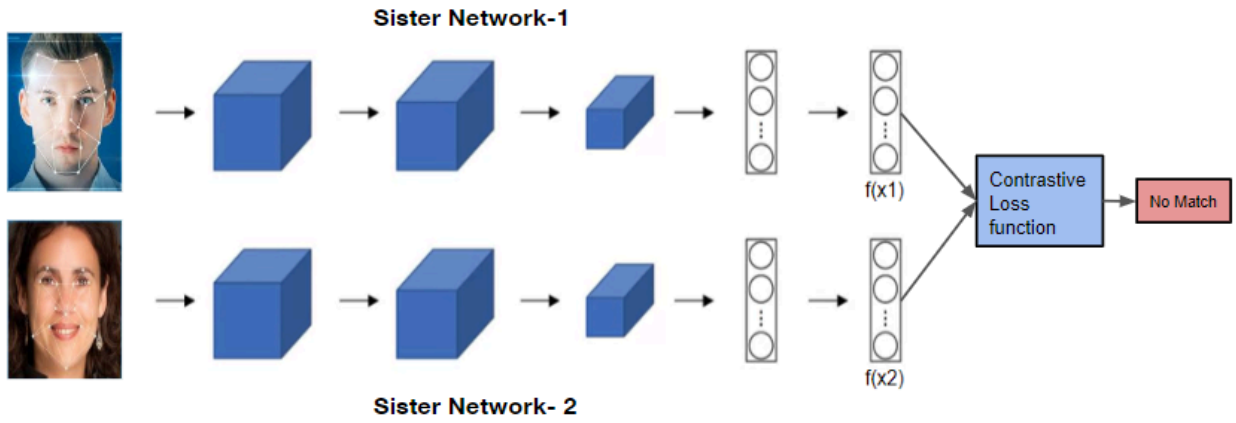
概念 : 把 refference feature 做成 kernel 去掃 query feature。將 output 結合後還原回 density map



用 Resnet50 的前三層，因此得到三組 Feature。
其中兩個 Resnet 的參數是一樣的，所以稱為Siamese Network (孿生網路)

$$Res_i^{query}, i \in \{1, 2, 3\}$$

$$Res_{i,j}^{ref}, i \in \{1, 2, 3\}, j \in \{1, 2, ..., k\}$$



J 張照片的同一個位置做 max pooling

$$Res_i^{ref} = max\_pool(Res_{i,j}^{ref}), i \in \{1, 2, 3\} \quad (3)$$

把 Res_ref 這張 feature 圖經過 max pooling 做成 r * r 的 kernel 做 conv

$$input = self\_attn(Res_i^{query}) \quad (4)$$

$$kernel = max\_pool(Res_i^{ref}, r) \quad (5)$$

$$M_i = Conv(input, kernel), i \in \{1, 2, 3\} \quad (6)$$

這個 kernel 不夠好，手腳會混再一起

會得到 M1, M2, M3

而 M1, M2, M3 有 k 個 channel ，因為 Refference images 有 k 張

這邊的 conv 是 1 * 1 conv ， sum 將 m1, m2, m3 變成 scalar

S = [sum_m1, sum_m2, sum_m3]

W = [a%, b%, c%]

$$S_i = Sum(Conv(M_i)), i \in \{1, 2, 3\} \tag{7}$$

$$W = Softmax(S), W \in R^3 \tag{8}$$

$$F = \sum_{i=1}^{3} W_i \times M_i, i \in \{1, 2, 3\}$$

bilinear upsampling.

Loss Function

$$L_E = \frac{1}{N}\sum_{i=1}^{N}\|P_i - GT_i\|^2,$$

SSIM loss

catching local pattern consistency

$$L_{SSIM} = 1 - \frac{1}{N}\sum SSIM(\mathbf{x}),$$

Experiments

1. 資料 :

將 COCO (Object Detection Dataset) 做 4 fold ，保留一個 fold 當作 testing

都挑 refference object 有五個以上的照片當作 query image, 加快訓練速度

再依照類別取 refference

refference 張數都取 5

生 density map : 把 box 的中心 標記為 mean，然後用 gaussian filter

the query image is randomly cropped to 256×256

reference images are resized to 64×64 with padding to keep the aspect ratio

2. 衡量標準

   SSIM

   - **亮度 (Luminance)**：比較兩張影像的平均亮度。
   - **對比度 (Contrast)**：比較兩張影像的對比度（通常用標準差來衡量）。
   - **結構 (Structure)**：比較兩張影像的結構資訊（通常用協方差來衡量，這反映了像素之間的相關性）。

   Predict ⇒ 用積分來算 density map 上的數量

   Cnt ⇒ 用積分來算 density map 上的數量

$$MAE = \frac{1}{M}\sum_{i=1}^{M}|Pred_i - Cnt_i| \tag{15}$$

$$MSE = \sqrt{\frac{1}{M}\sum_{i=1}^{M}|Pred_i - Cnt_i|^2} \tag{16}$$

code

https://github.com/SinicaGroup/Class-agnostic-Few-shot-Object-Counting

可以學習的地方

問題