Final Project

簡晟棋、李之曦、王韋翔

1. Abstract

在現實中會出現的異常檢測問題，除了local的structural anomaly之外，還有global的logical anomaly。由於現今多數異常檢測method都是針對structural anomaly，因此我們想結合VLM的方法，加強在兩種anomaly detection task上的的效果。

我們使用MVTec Loco AD的Dataset來訓練並評估我們的model。而我們利用了pre-trained model並結合SimpleNet、GCAD等的技術，來加強在兩種task上效果。

2. Introduction

VAND2.0 Challenge at CVPR competition要求參賽者在few shot learning的條件下訓練出能夠分辨出MVTec Loco AD Dataset中的structural anomaly與logical anomaly的model，並使用AUROC_score作為評分的標準。

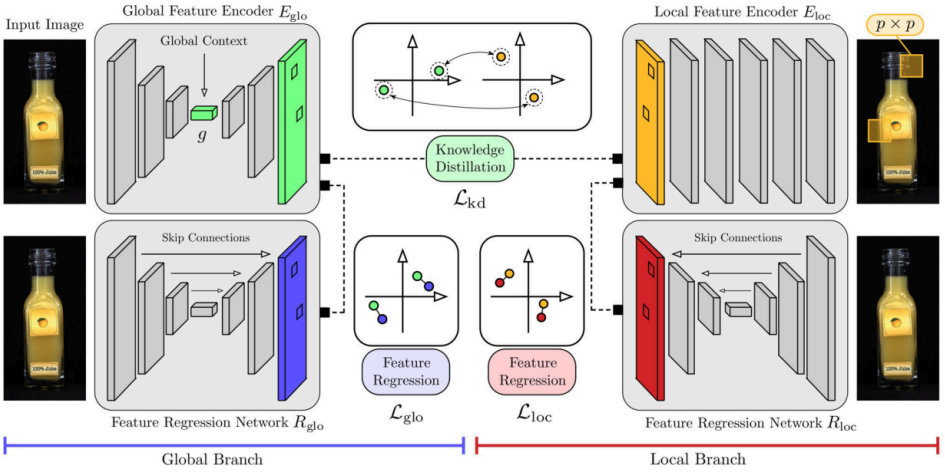
3. Related Works

SimpleNet[1][2](baseline):

我們參考了paper分別建構Extractor、Adaptor跟Discriminator，Extractor會透過pre-trained model將image encode得到features，接著透過few-shot learning過後的Adaptor得到adapted features，最後再加上noise，然後讓Discriminator判斷是否有被加過noise，藉此來學會如何偵測anomaly。預測時則會直接將adapted features直接輸入Discriminator，並藉由回傳結果判斷是否有異常。

GCAD[3](main method):

我們參考paper設計出Global Branch跟Local Branch，分別來處理logical anomaly跟structural anomaly。兩種Branch個別包含了一個feature encoder跟feature regressor，在預測時會根據encoder跟regressor的輸出結果判斷是否有anomaly。



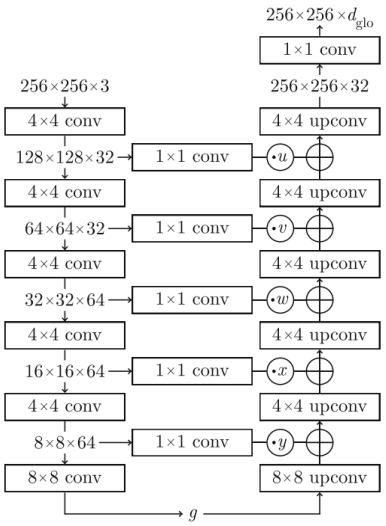
GCAD架構

4.Method

在訓練前，我們有對data做augmentation(RandomHorizontalFli、RandomVerticalFlip、RandomRotation(degree=30)、ColorJitter)，把training set數量變為5倍。

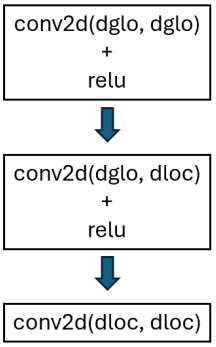
由於一般的model在structural anomaly已經有良好的表現，因此local encoder使用了pre-trained model(paper用的是resnet18，從中抽取128個參數，我們改成wide_resnet50_2，從中抽取1792個參數。此外，為了節省時間，paper中每個點代表經過pre-trained model的p*p方塊的中心點的值，我們改成將圖片分割成16*16的方格後經過pre-trained model後的結果)來幫助偵測。

而global encoder我們則參考了paper[3]設計出如下圖的model，upconV用ConvTranspose2d，u,v,w,x,y的參數在training時會從1開始陸續遞減，在1/5的epochs後全都歸0，testing時則都為0



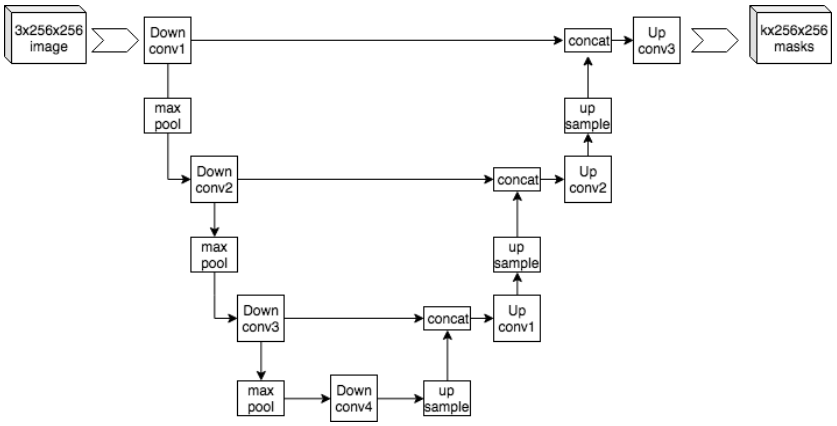
Global Feature Encoder架構

我們透過knowledge distillation(Upsample model)的技術，將local encoder的成果與global encoder的成果相比較。



knowledge distillation架構

在local與global的feature regressor方面，我們採用2種model實驗，1種是如paper中採用UNet[4]來幫助異常偵測。與下方示意圖不同的是實際共有5層，channel分別為[64,128.256,512,1024]。



Feature Regressor (U-net)架構，圖片來源:[5]

另一種則是基於原本的global encoder中使用的model結構跟UNet(將concat的部分換成add、捨棄maxpool)十分相似，因此我們嘗試了將該global encoder的model架構移植到global跟local的feature regressor上，使add weight全都保持在1，並增加channel數量至與原U-net數量一致。結果取得了相較於原本的設計更好的結果。

Loss function計算:local encoder與local regressor結果的誤差、global encoder與global regressor結果的誤差local encoder以及global encoder在knowledge distillation後的誤差的總和，如下圖。

$$\mathcal{L}_{\text{loc}}(I) = \|E_{\text{loc}}(I) - R_{\text{loc}}(I)\|_F^2$$

$$\mathcal{L}_{\text{kd}}(I) = \left\|E_{\text{loc}}(I) - U(E_{\text{glo}}(I))\right\|_F^2$$

$$\mathcal{L}_{\text{glo}}(I) = \left\|E_{\text{glo}}(I) - R_{\text{glo}}(I)\right\|_F^2$$

$$\mathcal{L}(I) = \frac{1}{d_{\text{loc}}}\mathcal{L}_{\text{kd}}(I) + \frac{1}{d_{\text{glo}}}\mathcal{L}_{\text{glo}}(I) + \frac{1}{d_{\text{loc}}}\mathcal{L}_{\text{loc}}(I)$$

anomaly score 計算:structural anomaly score 為local encoder與local regressor結果的誤差、logical anomaly score 為global encoder與global regressor結果的誤差，綜合結果為2者標準化後相加，如下圖。

- $A_{loc} = ||E_{loc}(J) - R_{loc}(J)||^2$
  - for detecting structural anomalies
- $A_{glo} = ||E_{glo}(J) - R_{glo}(J)||^2$
  - for detecting logical anomalies

$$A = \frac{A_{\text{loc}} - \mu_{\text{loc}}}{\sigma_{\text{loc}}} + \frac{A_{\text{glo}} - \mu_{\text{glo}}}{\sigma_{\text{glo}}}$$

5. Experiment
(Total, Structural(只有normal與Structural anomaly), Logical(只有normal與Logical anomaly)l)

AUROC	SimpleNet (baseline)	GCAD	GCAD(self)
breakfast_box	(0.59,0.61,0.57)	(0.63,0.64,0.64)	(0.63,0.66,0.63)
juice_bottle	(0.50,0.51,0.49)	(0.64,0.64,0.62)	(0.65,0.65,0.63)
pushpins	(0.56,0.65,0.49)	(0.55,0.64,0.46)	(0.64,0.75,0.55)
screw_bag	(0.52,0.50,0.54)	(0.58,0.59,0.59)	(0.59,0.61,0.57)
splicing_connectors	(0.49,0.47,0.50)	(0.61,0.61,0.59)	(0.51,0.51,0.50)
average	(0.53,0.55,0.52)	(0.60,0.60,0.58)	(0.60,0.64,0.60)

6. Discussion

我們設計的GCAD對於Structural和混合anomaly，在幾乎全部的class有最好的表現，某些class的Logical anomaly在原本的GCAD有較好的表現。splicing_connectors則是較為困難的class，相較於我們設計的GCAD，原本paper的model表現得更好。

由於硬體與時間原因，我們只有用100個epoch去訓練，更多epoch或許能得到更好的結果。

7. Conclusion

可以發現相較於structural anomaly，logical anomaly detection是相對更困難的任務。而相較於GCAD，SimpleNet只有考慮到Structural anomaly，因此做出來的結果相對較差。而比起原本的GCAD，我們改良過後的版本在這個問題上則有相對較好的平均結果，可能是因為比起UNet，paper中另外設計的model更擅長處理這個問題。

8. Reference

[1]https://openaccess.thecvf.com/content/CVPR2023/papers/Liu_SimpleNet_A_Simple_Network_for_Image_Anomaly_Detection_and_Localization_CVPR_2023_paper.pdf
[2]https://github.com/DonaldRR/SimpleNet
[3]https://link.springer.com/article/10.1007/s11263-022-01578-9
[4]https://github.com/jvanvugt/pytorch-unet
[5]https://zh.wikipedia.org/zh-tw/U-Net