ML project Details

Project Goal – Do visual object detection (SSD) then domain transfer using (CycleGAN), then do Domain Adaption in this domain. Implement another domain Adaptation.

SDD300 – 300X300 Input

1. Goal: Single shot detector is used for detection and localization of the object.
2. Datasets – The SDD implementation as proposed in the paper was implemented on pascalVOC dataset which is a fully annotated datasets on two things class and bounding box. The author decided to split the pascalVOC on trainval and test. The trainval is used for training and the test is used for testing the model.
3. SDD Input at the time of training

Input Image with Ground truth boxes and the object categories.

1. SDD Output obtained at test:

Images with bounding box and class scores.

1. SDD Architecture:

Baseline - VGG-16, Auxiliary Conv - Extra Feature Layers with provides higher-level feature maps, Prediction conv – locate and identify objects.

1. SDD Algorithms:
2. The Input Image is given to the network.
3. In the Auxiliary Conv, it will take features maps of different scales (7x7), (3x3) and draw grid in this feature maps.
4. Then we predefined boxes of various aspect ratio and scales per feature map location.
5. Then we predict the location (dx,dy,dw,dh) and the confidence scores.
6. The model is sum of two losses – localization and confidence loss.
7. Localization loss- These parameters include the offsets for the center point (cx, cy), width (w) and height (h) of the bounding box. The localization loss is smooth L1 loss, between the predicated box and the ground truth box parameters.
8. Confidence Loss – Which tells the presence of object in a default box (1-0). The closer it is to 1 the more confident the model is.The confidence Loss is the softmax loss over multiple class confidence. xij^p = {1,0}, is an indicator for matching i-th default box to the j-th ground truth box of category p.
9. Matching Strategy: It considers only positive matches and negative matches are discarded. When the predefined boundary box have an IoU greater than 0.5 with the ground truth box, it is positive.
10. Performance - **SSD300 achieves 74.3% mAP at 59 FPS**while**SSD500 achieves 76.9% mAP at 22 FPS.**
11. Output = Class , Score, Location = Apple , 0.92, [18,21,57,63]

1. SDD Training :

Give: Input Image with ground truth boxes and label for each object during training. At the time of training, we compare the predefined boxes with the ground truth boxes.

We select the one, which matches the highest confidences scores between positive and negatives 3:1.

1. SDD Testing: We give ground truth boxes along with the labels and see if it can do localization and detection. The object categories at the time of train and test is same.
2. SDD hyper parameters:
3. Optimizer: SGD,
4. LR: 10 ^-3 till 40k, next 10 ^-4 till 50k , next 10^-5 till 60k
5. momentum: 0.9,
6. Weight decay: 0.0005,
7. batch size: 32,
8. No dropout.

Our Project

1. SDD300 Training: In our project, we will take only the trainval split of pascalVOC2007 and pascalVOC2012 with the ground truth boxes and labels.
2. SDD Training Hyper parameters:
3. Lr for 40k : 10^-3 , Next 10k Lr: 10 ^-4 , Next 10k Lr: 10^-5,
4. Iterations : 60k
5. Batch Size-1
6. Epochs-1
7. Iterations – 60k
8. IoU = 0.45
9. Confidence score -0.01
10. Data Augmentation:
11. SDD Testing on Clipart: Are you testing with clipart annotation datasets?
12. Compare the score with the baseline of DA: After the end of testing, we need to check it with the baseline model.

CycleGAN

1. Goal -Translation of Image from the source Domain to the target Domain without paired images. It learns the mapping function between two domains (X, Y) without any paired examples.
2. Algorithm – There are two GAN. One mapping function source to target (G: X->Y) and other mapping function (F: Y->X) target domain to source domain.
3. Input for training: Just images without any annotations this is fully unsupervised. (Source Domain)
4. Output obtained at test: Just Domain Transfer images without any annotations.
5. Learning Rate: The cycleGAN learning rate that is used is 10x^-5
6. Discriminators – Discriminators are responsible for discriminating the images. (Classification model)
7. Generators - Generators are responsible for generation of fake images.

Oral

1. Explain how GAN works?
2. Explain how in CycleGAN works? How the images is being transferred into different domain?
3. Explain the overall loss of the model?

Project

Input – trainA : Pascal VOC2007/VOC2012 datasets

trainB : Clipart

Output - The images of pascalVOC2007/VOC2012 domain is obtain in clipart Domain.

Training – We are training our model to get transformed from trainA( source domain)to trainB (target domain) without any labels.

Testing – We are testing our model testA (source domain) to testB (target domain ), which gives the PascalVOC images in the clipart domains and we are saving this images to perform step3 of the project.

Hyper parameters –

1. lr -10^-5 ,
2. epochs-20 ,
3. iterations - ,
4. batch size- 1

Compare the score with the DT of the DA- After doing the domain transfer we compare the score with the DT of the DA on pascalVOC datasets

Domain Adaption

Goal - The goal of DA is to do detection and localization of images in variety of domains without the need to have instance level annotations in the target domain.

Source Domain – Instance Level annotations in source domain. (PascalVOC)

Target Domain – Image Level annotations (Clipart). The classes to be detected our subsets.

Instance Level Annotation- It is composed of label (object class of that instance) and bounding box (object location in that instance).

Image Level Annotation- Labels of object in each Image.

Algorithm: There are two steps, for Domain Adaptation

1. Domain Transfer – Domain Transfer is transferring the images of the source domain to that of the target domain and accompany with instance-level-annotations.
2. Pseudo Labelling – Pseudo Labelling is use in order to generate images with instance level annotations in the target domain.

Training Baseline : a. Train in the pascalVOC domain image, which got the annotations.

b. Images at the pascalVOC trainval in clipart domain is used for training, which got annotations.

Testing: Images in the target domain (clipart), which is split in train and test (1:1). Training images got no annotation, while test images got annotations. Clipart is splitted in 3 parts- train,test, traintest , train on clipart images,test after training, traintest – Don’t train on clipart just provide the annotation and test.

Hyper parameters: The hyper parameters used are same for SSD and CycleGAN.

Self-Supervised Learning for Domain Adaption

1. Goal- The main focus of this paper, i.e. can we incorporate self-supervision to learn domain adaptation.
2. Self-Supervision – There is two tasks- pretext task and the main task. Pretext task contains supervised samples and this should compensate for the lower number of manually needed samples for the main tasks.
3. Main Task – Semantic Segmentation.
4. Pretext Task- Object Recognition.
5. Source domain-[Xs,Ys] are labelled images and Target domain-[Xs] is unlabeled.
6. Result to obtain- We have a model, which is trained for semantic Segmentation in the target domain.
7. Main Task - E is encoder, which is responsible for feature extractor, S is decoder, E+S-> CNN semantic Segmentation. This CNN is trained end-end with our source domain labelled sample.
8. Pretext Task- E is encoder, which is shared with our main task, P, is the network for the model for the pretext Task. E+P-> CNN for the pretext task. The pretext task samples are automatically created by the target domain.
9. Training- We are jointly training our model both on the main task and on the pretext task. During the forward propagation, both the source and the target domain samples are propagated to the encoders. After that, the losses of the main-task and the pretext tasks are computed and they are back propagated. The encoder is trained on both the main task and the pretext task which means both on the source domain and on the target domain invariant feature represenatations.
10. Testing- We feed the target domain images to the decoder which is unlabeled and we ask them to make the predictions.
11. How we are actually automatically generating samples for the pretext tasks and do the object recognitions (image rotation prediction model)? – We are taking images from the target domains which are unlabeled and we are performing 2D transformation rotations (0, 90,180,270) this is how the automatically samples are generated. Now do object recognition, at first the input image is provided to the encoder(E), the encoder is extracting feature maps and providing to the P as input and as output P is providing probability distributions overall all the possible geometric transformations.
12. What the pretext task is actually helping? – The pretext task is helping the encoder to learn domain invariant features about the target domain.

Our Project

Source Domain – Pascal VOC

Target Domain – Clipart in unlabeled.

Main Task- Object Detection

Pretext Task – Image rotation

Baseline-

Steps for the DA Project – Pixel Level DA

1. Source Domain – PascalVOC 2007 /2012 (natural images)
2. Target Domain- Clipart (cartoon )
3. At first we train out model on SSD on pascalVOC . (Give the boundary box and the labels)
4. Then we test our model on clipart, which is a different domain. (Give the boundary box and the labels).
5. Then we compare the test result with the baseline and we see that SSD is able to do object detection and localization without being trained on that domain. The result is poor as the training and the testing is done on different domain.
6. Then we do domain transfer (DT) using cycleGAN after that the images of pascalVOC looks like that of clipart domain.
7. Then we perform training for each epoch 10K iterations, this time for the training we are using the pascaLVOC transferred domain images.
8. Then we test our model on our target domain (clipart).The target domain is split into two parts at the time of training we don’t provide our model with bounding box, but at the time of testing we do. (train,test,traintest)
9. Next we compare our result with the DT.

Questions

1. The images, which is obtained after domain transfer, do not have any annotations as CycleGAN do not perform annotations how we are able to use SSD on this.

We are able to use this because remember only the domain is transferred not the locations of the objects in the images so at the time for performing training on SDD on pascalVOC dataset we were having the annotation, we are getting instance level annotations from that set.

1. Why we need to compare the test on clipart with the baseline model?

This step is done to check if the SDD is performing well on a different domain on which it is not trained for, i.e. clipart.

1. Is you do training on clipart too before testing it on step 1?

No you train your model only on the source domain.

1. Is all the clipart images annotated ?

No all the images here are not annotated. (check)

1. What do we do after domain transfer (pascalvoc->clipart domain)?

After doing the domain transfer we have the images of the PascalVOC in clipart domain. We are doing again a training of this images (Training images of pascalvoc on clipart domain with annotation) and then testing it on the clipart images.

Hence now we see the maP (mean Average Precision) to increase the performances as the two domains of training and testing our same.

Oral

1. Can you explain how SDD helps in object detection?
2. Can you explain the various loss function used by the author?
3. Can you justify why you have decided to use those hyper parameters?
4. What other object detection techniques do you know?
5. Can you explain the difference between Faster-RNN and YOLO?