Object Recognition
Practical Sessions

Meysam Madadi
meysam.madadi@gmail.com

Date	Theory title	Teacher	Practical session
20-02 - 2024	Presentation and CNN basics	Sergio	P1 intro
27-02 - 2024	CNN and GNN architectures	Meysam	-
05-03 - 2024	Object detection and segmentation	Meysam	P1 Q&A - P2 intro
12-03 - 2024	Human pose estimation	Meysam	P1 presentation
19-03 - 2024	Human behaviour understanding	Sergio	-
26-03 - 2024	Easter holidays	-	
02-04 - 2024	INVITED TALK		P2 Q&A - P3 intro
09-04 - 2024	Exams week	-	
16-04 - 2024	Master seminar	-	
23-04 - 2024	Recurrent models and transformers	Meysam	P2 presentation
30-04 - 2024	Presentation I	Sergio	
07-05 - 2024	Presentation II	Sergio	
14-05 - 2024	Generative models	Meysam	P3 Q&A
21-05 - 2024	MONDAY schedule	-	
28-05 - 2024	Exam	Sergio	P3 presentation

7 sessions
3 blocks

Deliverables

1. CNN architecture design, deadline 11/03/2023 23:59
2. Fashion parsing (segmentation), deadline 08/04/2023 23:59
3. Body and clothes depth estimation, deadline 20/05/2023 23:59

Notes:

- The tasks must be done in the groups of **4** students. One student in each group must send me the name of his/her mates in the group in one week. If I do not receive this email, I will assign students to the groups randomly.
- Deliverable is a zip file including the code, report pdf and presentation slides.
- The report must fit within the **10** page limit using the font **calibri** with size **11**. Extra pages will be penalized.
- Submissions must be done through virtual campus.
- Submissions after the deadline will be penalized.
- The solutions and results must be presented by the members of each group in 10 minutes on the defined days,
- Deliverables are evaluated based on the quality of the report and presentation, and the number of tasks done. Erroneous results without proper justification won't be taken into account.

Deliverable 1 - CNN architecture design

- The goal is to learn how to design and train CNN networks to maximize the performance,
- The task is the multi-label classification on the **Pascal VOC 2012** dataset,
- You can use the given code in the practical session as your base code.

Deliverable 1 - Tasks

1- Analyze the dataset for the distribution of the labels by:

a. Counting the objects;
b. Counting the images, that is no matter how many objects of the same label appear on the image, all counts 1;
c. Mean and variance of the area ratio (dividing object area to image area) of objects per label. You can use object bounding box to compute the area.

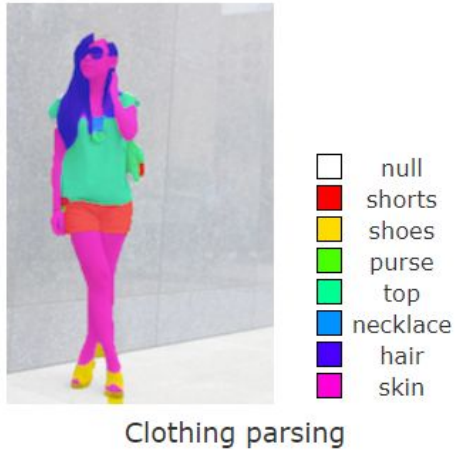
Deliverable 1 - Tasks

2- Train and evaluate the ResNet50V2, EfficientNetV2B0 and ConvNeXtTiny from scratch vs pretrained on imagenet using the default hyperparameters and classifier head given in the practical notebooks. Note that the training and evaluation must be done on the given train/test txt files. Then select the best performing network and do the following tasks incrementally:

a. Tune the batch size using this list (as far as memory allows) [16, 32, 64, 128].
b. Develop two data augmentation algorithms of your choice along with the default ones in the given notebook.
c. The dataset is unbalanced. Develop two algorithms of your choice to have a better balance among labels during the training without sacrificing the accuracy.
d. Redesign the classifier head to have better results.
e. Increase the number of epochs as much as needed to maximize the test set performance.

Write the report by explaining, comparing and analysing the results.

Deliverable 2 - fashion parsing



What do you need to do in this task?

- Fashion semantic segmentation

On which dataset?

- Fashionpedia: https://fashionpedia.github.io/home/

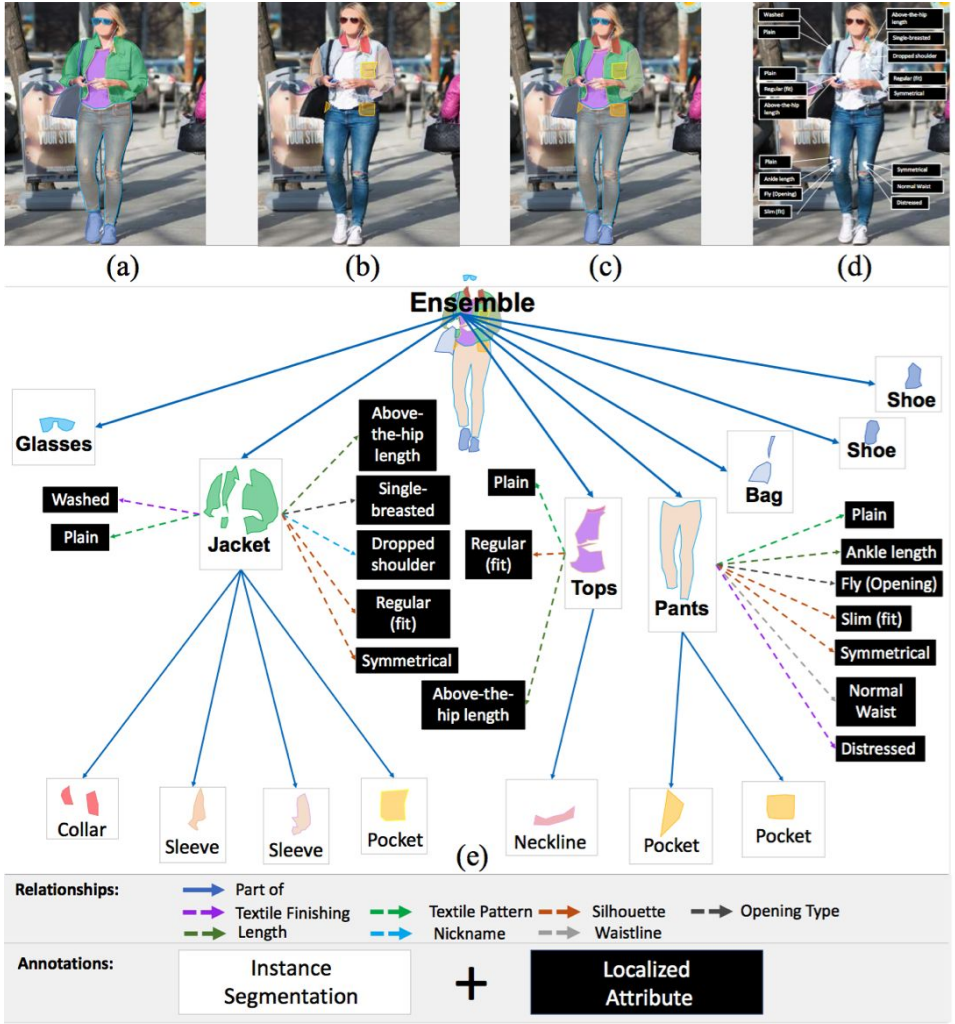
Using which code base?

- MMSegmentation: https://github.com/open-mmlab/mmsegmentation
- Or any code you find more convenient

Fashion parsing

Hierarchical segmentation and attribute detection

- We do not consider item parts and attributes,
- Only the person in the middle has been annotated.
- We use the validation set to report the results.



Deliverable 2 - Tasks

1- Analyze the dataset for the distribution of the labels by:

- Counting the number of images per label,
- Mean and variance for the pixel ratio per label:
  - (# of pixels of label j in image i) / (# of pixels in image i),
  - (# of pixels of label j in image i) / (# of **foreground** pixels in image i) ====> in this case we exclude the background

Deliverable 2 - Tasks

2- Choose three segmentation networks (justification is required) from the MMSegmentation implementations and finetune them on the fashionpedia dataset, as far as your resources allows, as following:

- Use the default learning rate as suggested by MMSegmentation,
- Tune batch size based on your resources,
- Apply the basic data augmentation like rotation, scaling, cropping, etc,
- Train based on 192 vs 384 px image resolution. Is there any relationship between the resolution and accuracy of each label?

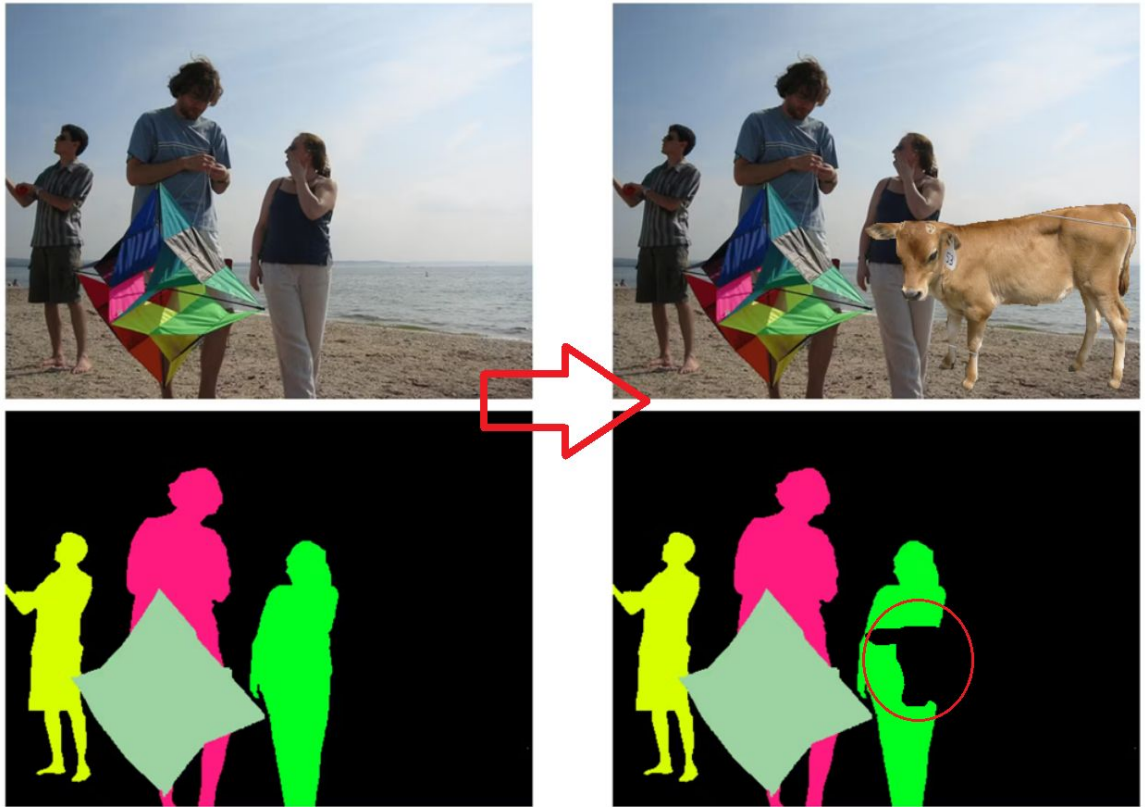
Select the best performing network and resolution and do the following:

- Discard the large and overrepresented labels and train the network with the same hyperparameters as above. Does this make any difference to the results of small objects?

Deliverable 2 - Tasks

3- Implement the following data augmentation and train the best performing network:

- Using the instance segmentation annotations of PASCAL VOC dataset, copy object instances randomly and paste them in random locations on the target people in fashionpedia images while updating the ground truth segments.



Deliverable 2 - Tasks

4- Without implementing any code or training, just discuss two possible ideas to improve the results like network modification, higher order relationships as extra heads, losses, leveraging unlabeled data, etc

NOTES for writing the report:

- Justify your choices,
- Thoroughly discuss the results both qualitatively and quantitatively,
- Report the results using accuracy and mDice metrics excluding the background,
- Your best performing network must show at least 20% mDice