M5 Project: Object Detection and Segmentation
Team 3

Hyperparameters Selection

Keyao Li (rylesssi@gmail.com)
Marc Orós Casañas (marc.oros@e-campus.uab.cat)
Daniel Fuentes(daanyfuentes@gmail.com)

Outline

1. Introduction
2. Content
    1. Learning Rate
    2. Batch size
    3. Scheduler
    4. Top K Train
3. Findings & Conclusions
4. Q&A

Introduction

We will introduce different hyperparameters to understand their impact on the results.
The hyperparameters we will analyze are as follows:

- ❑ **Learning rate** - [0.0025, 0.0001, 0.00025]
- ❑ **Batch** - [4, 8]
- ❑ **Scheduler** - [WarmupMultiStepLR, WarmupCosineLR]
- ❑ **Top K Train** - [6000, 9000, 12000]

In order to conduct different experiments, the fixed parameters are as follows:

CNN	R50-FPN
Dataset	KITTI-MOTS
Learning Rate	0.0025
Batch	4
Scheduler	WarmupMultiStepLR
Top K Train	6000

Explore and analyze the impact of different hyperparameters — Learning rate

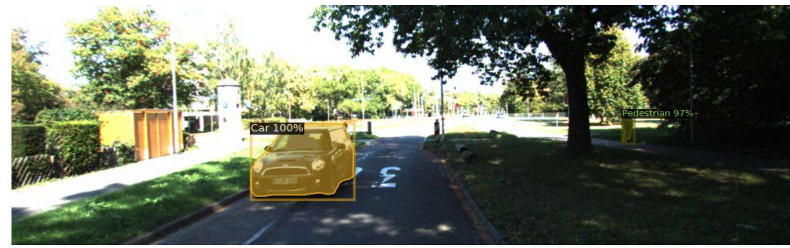
The learning rate controls how quickly the model is adapted to the problem.

Smaller learning rates require more training epochs given the smaller changes made to the weights each update, whereas larger learning rates result in rapid changes and require fewer training epochs.

Explore and analyze the impact of different hyperparameters — Learning rate

LR 0.0025

LR 0.001

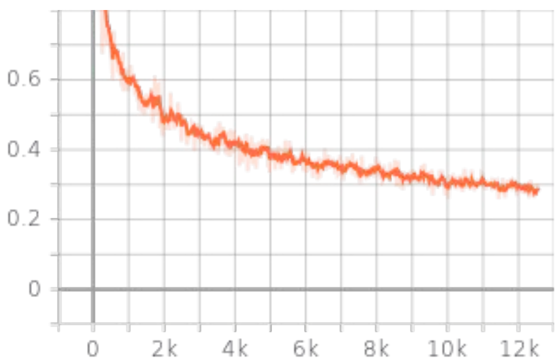


LR 0.00025

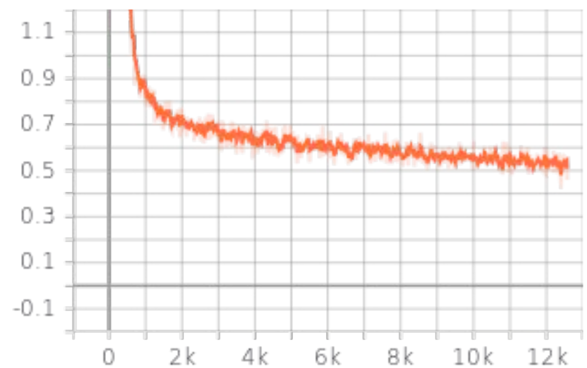
In this results we can see perfectly the effects of different learnings rates, in small lr we can not detect all in intermediate lr is perfectly detected and segmented and finally with big lr we are facing false positives.

Explore and analyze the impact of different hyperparameters - Learning Rate

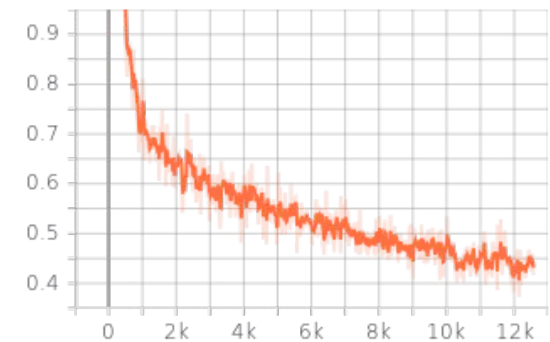
LR 0.0025



LR 0.001



LR 0.00025



AP	AP50	AP75	APs	APm	APl
Learning Rate - 0.0025- Time training 6972s					
39.3	55.3	44.8	25.3	53.6	26.8
33.0	52.3	33.8	16.9	45.1	38.4
Learning Rate - 0.0001- Time training 8494s					
44.4	57.9	45.8	28.2	52.2	31.2
33.8	55.1	35.2	19.0	46.1	42.2
Learning Rate - 0.00025- Time training 8574s					
39.4	56.3	45.1	27.1	51.6	30.1
33.0	53.7	35.2	18.4	45.6	37.7

As we can see the direct effect of learning rate is in the loss curve.
Big lr produces big slopes than small lr.
We can confirm the qualitative results. we have the bests AP's with learning 0.0001.

Also one thing to take into account is that with small learning rates we can arrive to the max iteration as this is the case with 0.00025.
This is important because sometime touch one parameter is limited by other one.

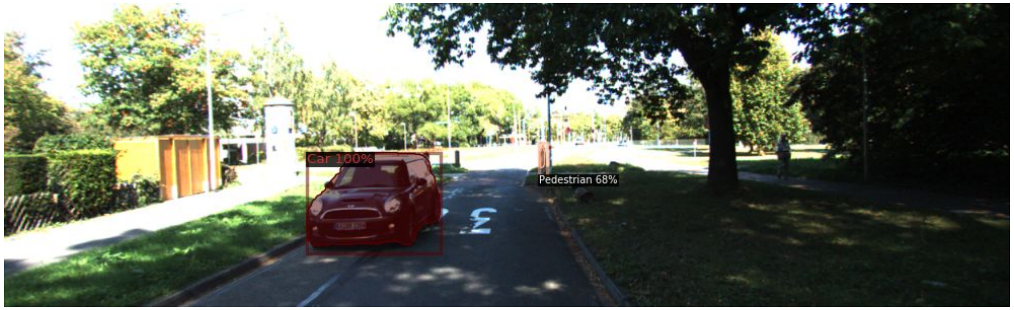
Explore and analyze the impact of different hyperparameters - Batch Size

The batch size is the amount of images that our neural network processes at a time. A large batch size will mean that our network will train more quickly, but a batch size that is too large might make it so our network doesn't generalize. Additionally, our batch size might be limited by the amount of memory on our GPU.

Explore and analyze the impact of different hyperparameters - Batch Size

Batch 4

Batch 8



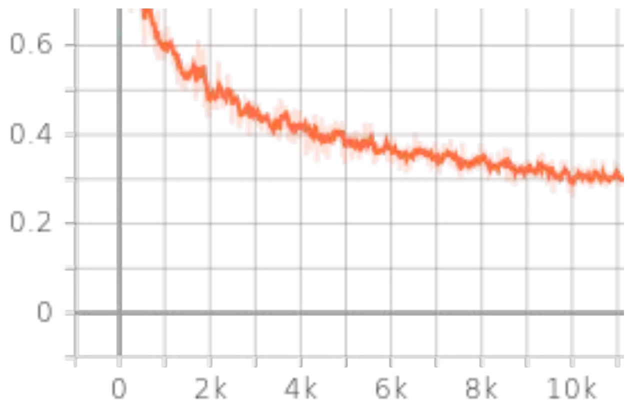
We only tested with a batch size of 4 and 8, as the GPU didn't allow a bigger batch size because it didn't have enough memory.

As we can see, the results are very close, with the bigger batch size creating a false positive, but the mask around the car is very close.

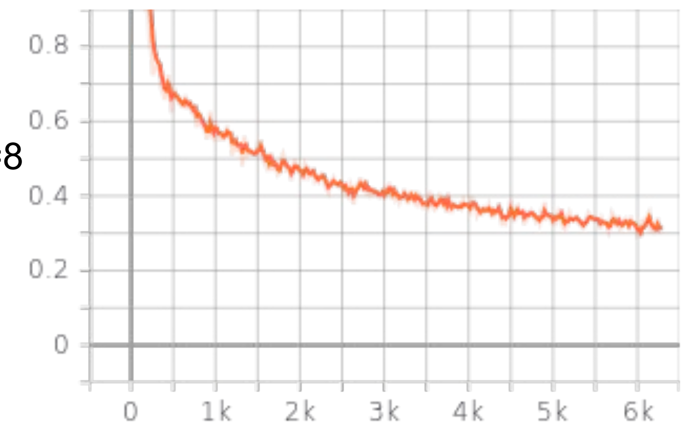
While not visible here, increasing the batch size can allow us to get better results in the same amount of time.

Explore and analyze the impact of different hyperparameters - Batch Size

Batch size=4



Batch size=8



AP	AP50	AP75	APs	APm	APl
Batch - 4 - Time training 6972s					
39.3	55.3	44.8	25.3	53.6	26.8
33.0	52.3	33.8	16.9	45.1	38.4
Batch - 8 - Time training 5383s					
39.9	56.0	45.5	25.9	53.5	32.4
34.2	53.4	36.2	18.0	46.9	38.3
Batch - 16 - No Result					

As we can see in the graph, we can iterate over the same number of images in significantly less time. Additionally, at least in our case, increasing the batch size improves the results of our model.

Explore and analyze the impact of different hyperparameters - Scheduler

The scheduler is used to adapt the learning rate throughout the learning process of our network. A better scheduler can make it so we get better results for a set number of iterations by picking the most optimal learning rate.
For our testing we tried the WarmupMultiStepLR and WarmupCosineLR schedulers.

Explore and analyze the impact of different hyperparameters - Scheduler

WarmupMultiStepLR

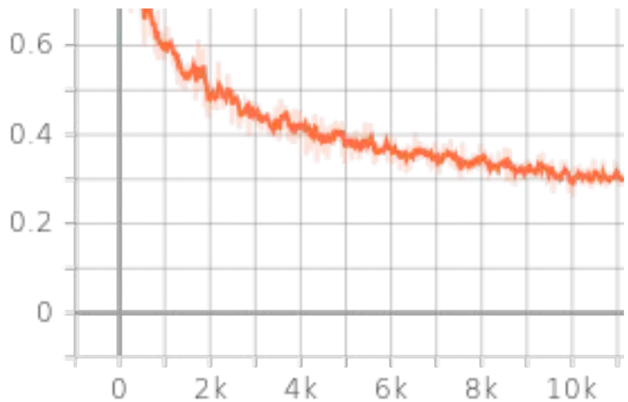
WarmupCosineLR

As we can see in the picture, the benefit of using a better scheduler is not immediately noticeable, as the gains of using a better scheduler are relatively small.

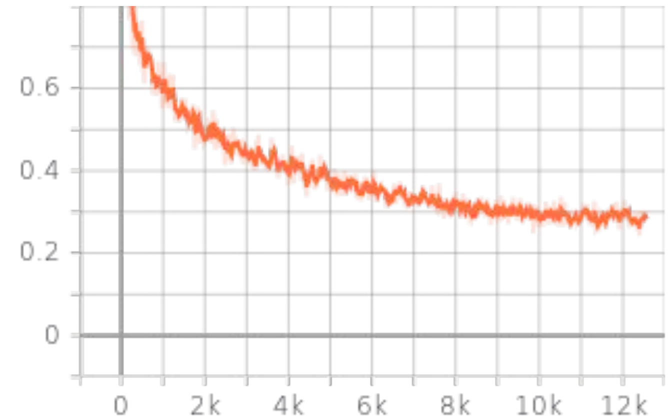
Explore and analyze the impact of different hyperparameters

Scheduler

WarmupMultiStepLR



WarmupCosineLR

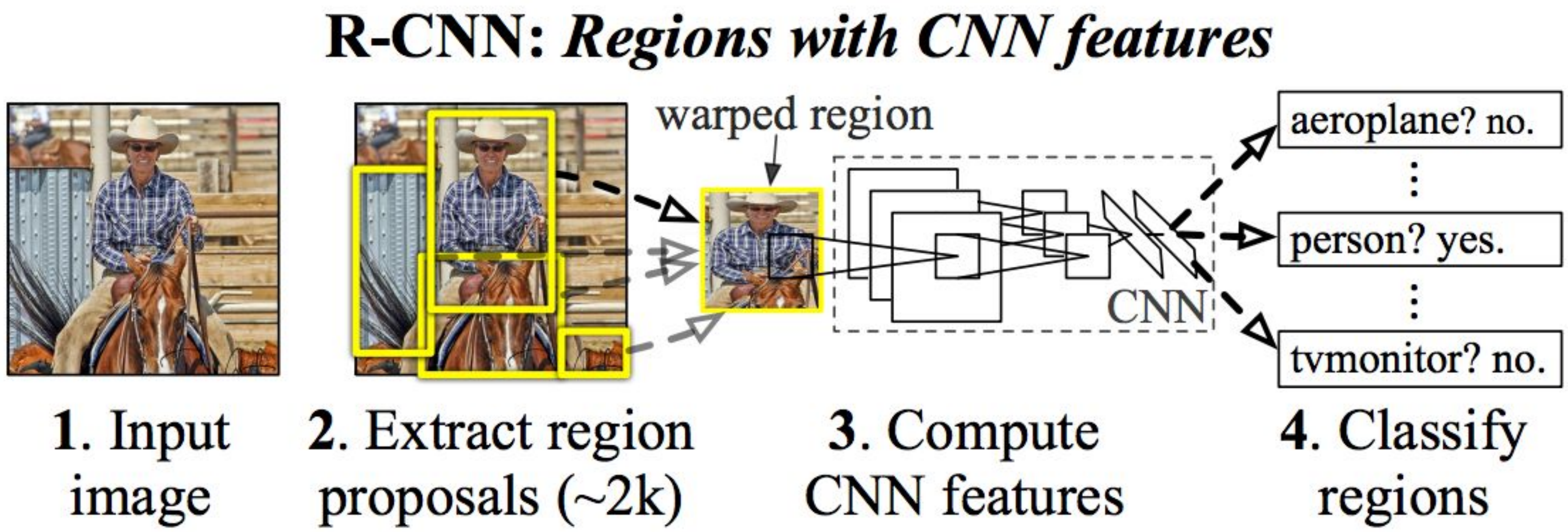


AP	AP50	AP75	APs	APm	APl
Scheduler WarmupMultiStepLR - Time training 6972s					
39.3	55.3	44.8	25.3	53.6	26.8
33.0	52.3	33.8	16.9	45.1	38.4
Scheduler WarmupCosineLR - Time training 5383s					
40.1	55.8	45.4	26.6	53.3	29.5
34.0	53.4	35.4	17.7	46.5	36.3

As we can see the main difference in the results is the amount of iterations needed to obtain the same results. In this case we can see that WarmupCosineLR gives us slightly better results than WarmupMultiStepLR

Explore and analyze the impact of different hyperparameters - Top K Train

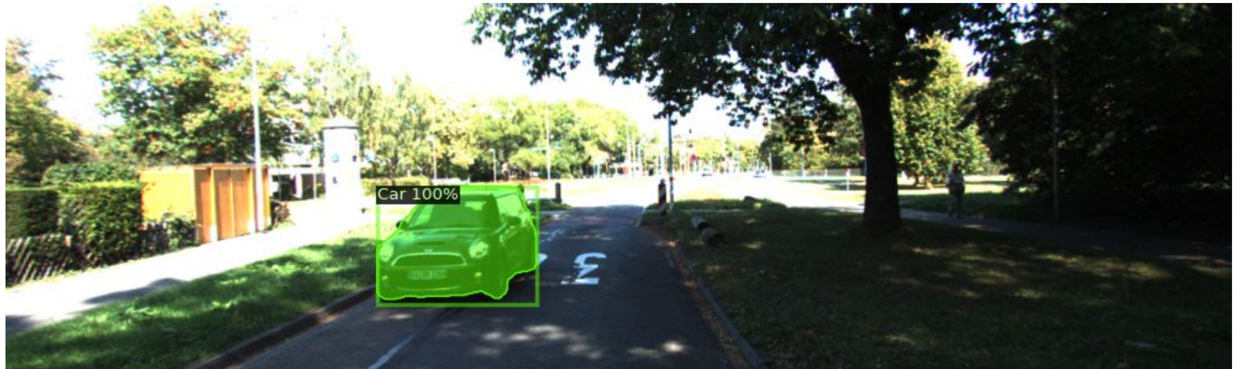
We use selective search algorithm for object detection to generate region proposals.



Why is important choose a correct value?

Explore and analyze the impact of different hyperparameters - Top K Train

Top K train 6000

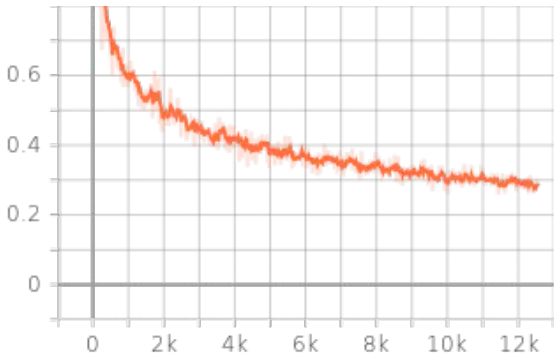


Top K train 9000

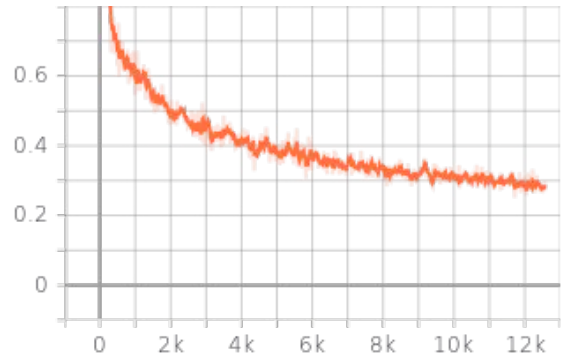
Top K train 12000

Explore and analyze the impact of different hyperparameters - Top K Train

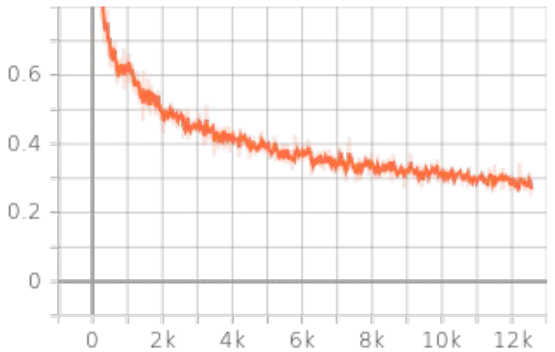
Top K train 6000



Top K train 9000



Top K train 12000



AP	AP50	AP75	APs	APm	APl
Top K Training - 6000 - Time training 6972s					
39.3	55.3	44.8	25.3	53.6	26.8
33.0	52.3	33.8	16.9	45.1	38.4
Top K Training - 9000 - Time training 10383s					
39.9	56.0	45.5	25.5	53.5	32.4
34.2	53.4	36.2	18.0	46.9	38.3
Top K Training - 12000 - Time training 12485s					
40.4	57.1	46.5	27.3	53.6	33.8
34.0	54.4	36.4	18.5	47.1	39.2

Explore and analyze the impact of different hyperparameters - Top K Train

How to select the correct value:

- Trade off between performance and results
- Metrics decay with less proposals, is necessary to find this inflexion point
- To much proposals can produce wrong matches, between detections or pixels in segmentation, we can have as a point of reference the number of false positives

AP	AP50	AP75	APs	APm	APl
Top K Training - 6000 - Time training 6972s					
39.3	55.3	44.8	25.3	53.6	26.8
33.0	52.3	33.8	16.9	45.1	38.4
Top K Training - 9000 - Time training 10383s					
39.9	56.0	45.5	25.5	53.5	32.4
34.2	53.4	36.2	18.0	46.9	38.3
Top K Training - 12000 - Time training 12485s					
40.4	57.1	46.5	27.3	53.6	33.8
34.0	54.4	36.4	18.5	47.1	39.2

Finding & Conclusions

- The correct selection of parameters plays a crucial role in the final results
- Sometimes we need to play with the trade off performance/results
- Is necessary to be careful with dependencies between hyperparameters.
- In our opinion this lab was the most relevant and interesting to implement


thank you
gracias
danke
謝謝
merci
obrigado
teşekkür ederim
go raibh maith agat
ngiyabonga
bedankt
dziękuję
sukriya
kop khun krap
terima kasih
mochchakkeram
спасибо
dank je
grazie
arigatô
dakujem
takk
tapadh leat
sagolun
hvala
ευχαριστώ
감사합니다
ありがとう
мерси
рахмат
Баярлалаа
хвала

Questions ?