Object detection, segmentation and pose estimation for mobile devices

Vitaliy Bulygin, Samsung R&D Institute



















Solution overview (for wide range CV tasks)



Solution overview (for wide range CV tasks)



Solution overview (for wide range CV tasks)



Input resolution



Model size trick



Model size trick



Model size trick



Connection between backbone and head



Deformable and transpose convolution



Task output



Task output



Task Output : Object Detector



Task Output : Object Detector



Task Output : Pose Estimation



Task Output : Semantic segmentation



Segmentation + Object Detector multi-task



Segmentation + Object Detector multi-task results

MobileNet v2 backbone, device : Samsung S10 Exynos

	mAP	mIOU	inference (ms)
segm only	_	18.6	
OD only	21.1	-	17 ms
OD + segm	22.0	19.1	19 ms

MS COCO 80 classes



Face part segm: detect head + segm



Segmentation + Object Detector multi-task results

MobileNet v2 backbone, device : Samsung S10 Exynos				
	mAP	mIOU	inference (ms)	
segm only	-	18.6		
OD only	21.1	-	$17 \mathrm{ms}$	
OD + segm	22.0	19.1	19 ms	



Nikita Dvornik et al "BlitzNet: a real time ...", 2017



Kaiming He et al "Mask R-CNN", 2017



segm+ OD gives better accuracy than OD alone

Segmentation and pose estimation result

MobileNet v2 backbone, device : Samsung S10 Exynos - 17 ms for PE and segm, 10 ms for lite OD







ImageNet pre-train



ImageNet pre-train



big batch per gpu + batch norm layers Rui Zhu etc "ScratchDet: Training ...", 2019 long warm-up for large learning rate Leslie Smith "A Disciplined Approach ...", 2018 Better optimizer without pre-train training accuracy ↑ time↓↓

Sharp vs Flat local minima



Learning rate schedule: common approach



Learning rate schedule: common approach



Optimizers: literature analysis

- 1) increase learning rate as much as possible
- 2) get the best on plateau
- 3) decrease learning rate for small accuracy increment



- Leslie Smith "Super-Convergence: Very Fast Training of Neural Networks Using Large Learning Rates", 2017
- Mark Schmidt and Nicolas Le Roux. "Fast convergence of stochastic gradient descent under a strong growth condition".
- Yuanzhi Li, et al "Towards Explaining the Regularization Effect of Initial Large Learning Rate in Training Neural Networks"
- Rui Zhu, et al "ScratchDet: Training Single-Shot Object Detectors from Scratch"
- Leslie N. Smith "Cyclical Learning Rates for Training Neural Networks"
- Shirish Keskar et al. "On Large-Batch Training for Deep Learning...", 2017.



- 2)get the best on plateau
- decrease learning rate for small accuracy increment





step







Conclusions

- It is better to write your own detector
- Real-time mobile NN could has only low-resolution input
- Model size, accuracy and speed could be tuned
- There is a trick to greatly reduce model size with only a bit accuracy drop
- It is better to avoid transformed and transpose convolutions
- Flat extremums gives higher accuracy on test set
- Large batch size and learning rate lead to flat minima
- It is possible to quickly and simply add CV new task