PyTorch YOLOv3 Object Detection for Vehicle Identification

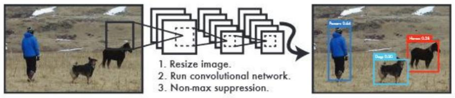
Tesa Ho, tesaho@stanford.edu; Mohith Ravendra, mohithr@stanford.edu

CS 230 Spring 2019

https://youtu.be/1o0FantqmPM

Goal

- Utilize transfer learning to train a YOLOv3 for vehicle detection.
- Avoid hand labelling video images by training on a combined set of stock car images (Stanford cars dataset) and real world video images (NEXAR dataset).



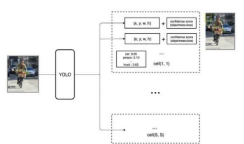
Data

- NEXET images (bottom) are different quality, lighting conditions, and perspective from Stanford car images (top).
- 9 vehicle classes.

	Train				Validation	
	Stanford	Nexet	Total	%	Nexet	%
sedan	4,851	754	5,605	58.5%	247	52.9%
hatchback	554	53	607	6.3%	17	3.6%
bus	0	60	60	0.6%	19	4.1%
pickup	593	92	685	7.2%	30	6.4%
minibus	0	0	0	0.0%	0	0.0%
van	541	248	789	8.2%	81	17.3%
truck	0	102	102	1.1%	33	7.1%
motorcycle	0	0	0	0.0%	0	0.0%
suv	1,605	123	1,728	18.0%	40	8.6%
Total	**8,144**	**1,432**	**9,576**	**100.0%**	**467**	**100.0%**

YOLOv3 – DarkNet53

- Pre-trained on ImageNet.
- Each image padded and resized to 416 x 416



	Type	Filters	Size	Output
	Convolutional	32	3 × 3	256 × 256
	Convolutional	64	3 × 3 / 2	128 × 128
1×	Convolutional	32	1 × 1	
	Convolutional	64	3 × 3	
	Residual			128 × 128
	Convolutional	128	3 × 3 / 2	64 × 64
2×	Convolutional	64	1 × 1	
	Convolutional	128	3 × 3	
	Residual			64 × 64
	Convolutional	256	3 × 3 / 2	32 × 32
8×	Convolutional	128	1 × 1	
	Convolutional	256	3 × 3	
	Residual			32 × 32
	Convolutional	512	3 × 3 / 2	16 × 16
8×	Convolutional	256	1 × 1	
	Convolutional	512	3 × 3	
	Residual			16 × 16
	Convolutional	1024	3 × 3 / 2	8 × 8
4×	Convolutional	512	1 × 1	
	Convolutional	1024	3 × 3	
	Residual			8 × 8
	Avgpool		Global	
	Connected		1000	
	Softmax			

Loss Function

- Sum squared error of prediction and ground truth.
- Composed of 3 losses:
  1) classification loss
  2) localization loss (predicted box and ground truth errors)
  3) confidence loss (objectness of the box)

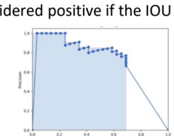
$$\lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{\text{obj}} \left[(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2 \right]$$
$$+ \lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{\text{obj}} \left[\left(\sqrt{w_i} - \sqrt{\hat{w}_i} \right)^2 + \left(\sqrt{h_i} - \sqrt{\hat{h}_i} \right)^2 \right]$$
$$+ \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{\text{obj}} \left(C_i - \hat{C}_i \right)^2$$
$$+ \lambda_{\text{noobj}} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{\text{noobj}} \left(C_i - \hat{C}_i \right)^2$$
$$+ \sum_{i=0}^{S^2} \mathbb{1}_{i}^{\text{obj}} \sum_{c \in \text{classes}} (p_i(c) - \hat{p}_i(c))^2$$

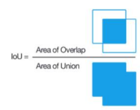
Evaluation

- The model was evaluated using the mean average precision (mAP) metric.
- Mean average precision is the average precision (AP) per class.

$$\mathrm{AP} = \Sigma \left(r_{n+1} - r_n \right) p_{interp}(r_{n+1})$$
$$p_{interp}(r_{n+1}) = \max_{\tilde{r} \ge r_{n+1}} p(\tilde{r})$$

- A prediction is considered positive if the IOU score >= 0.5.
- AP is also the area under the precision recall curve





Hyperparameter Search

- Learning rate: [0.00001, 0.0001 0.0005]
- Confidence threshold: [0.01, 0.05, 0.10]
- Non-maximal threshold: [0.30, 0.50, 0.80]
- No data augmentation
- Max epochs = 75

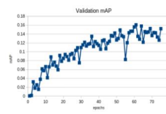
Table 1. Model parameters

Batch normalization	Yes
Batch size	6
Multi-scale training	Yes
Momentum	0.9
Decay parameter	0.0005
Learning rate	0.0001
Confidence threshold	0.05
NMS threshold	0.5
IOU threshold	0.5

Results

- Total mAP @ IOU_0.50 = 0. 1607
- Sedan achieved a class mAP of 0.5040 @ IOU_0.50

	mAP				
	IOU_0.005	IOU_0.25	IOU_0.50	IOU_0.75	IOU_0.95
sedan	0.6125	0.5916	0.5040	0.1685	0.0000
hatchback	0.0756	0.1406	0.1406	0.0867	0.0028
bus	0.0000	0.0000	0.0000	0.0000	0.0000
pickup	0.0456	0.0542	0.0617	0.0063	0.0000
minibus	0.0000	0.0000	0.0000	0.0000	0.0000
van	0.1916	0.1855	0.1937	0.1429	0.0000
truck	0.1088	0.0971	0.0900	0.0387	0.0000
motorcycle	0.0000	0.0000	0.0000	0.0000	0.0000
suv	0.1353	0.1353	0.1353	0.0813	0.0000
total	0.1670	0.1720	0.1607	0.0749	0.0004



Error Analysis

1) Poor lighting conditions
2) No training examples for some classes
3) Perspective issues
4) Poor visibility

Future: Data Augmentation

- Augment minority classes to address issues from error analysis.
- Flipping, scaling, brightness variation, perspective transform, image sharpening.

References

J. Redmon, A. Farhadi. “YOLOv3: An Incremental Improvement”, University of Washington. 2018.
J. Hui. “Real-time Object Detection with Yolo, YOLOv2, and now YOLOv3”. 2018.
J. Sang, Z. Wu, P. Guo, H. Hu, H. Xiang, Q. Zhang, B. Cai. “An Improved YOLOv2 for Vehicle Detection”. Sensors December 2018.