

Open Images 2019 - Visual
Relationship
3rd place sharing

by Very Random Team

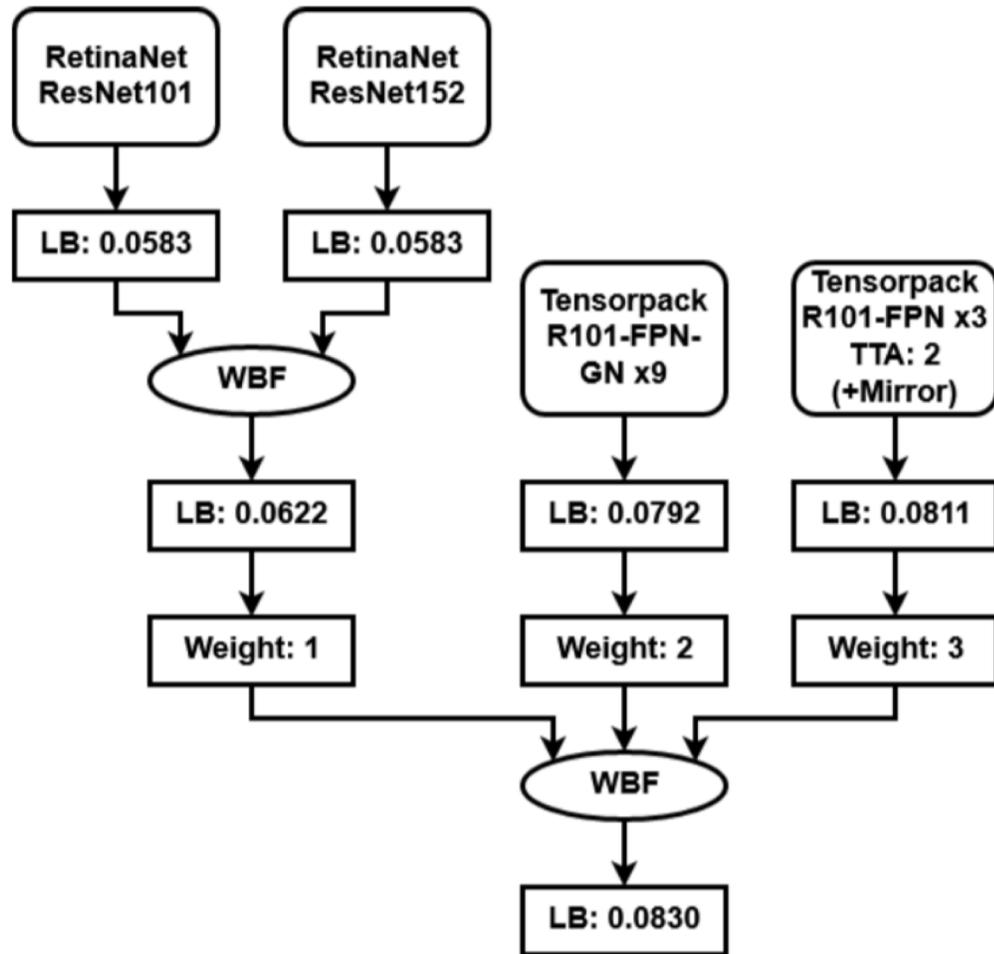
Solution overview

- Bounding box models (ZFTurbo, Weimin)
- Attribute model (ZFTurbo, Weimin, Authur)
- Relationship model (Vlad)
- Weighted Box Fusion (ZFTurbo, Weimin)

Bounding box models

- Faster R CNNs, with two types of training (using OD training data):
 - Starts with uniform sampling on images -> once converged, switches to sub-trains based on label counts: top 150, 150 – 300, 300 – 500
 - Starts with uniform sampling on images -> once converged, switches to class-balance sampling
- Used frameworks of *tensorpack*, *mmdetection*, *tensorflow* (*inception_resnet_v2*)
- Models weights initialized from ImageNet or COCO
- *Ensembled with Weighted Box Fusion (WBF)*
 - 0.46 (inception_resnet_v2) + 0.51 ~ 0.56 (tp+mmdetect, 5 models) -> 0.63 (pub)

Attribute model



- Used 42 'is' labels, such as *Handbag is Textile*, *Ski is Wooden*, etc
- Trained the models as typical OD model
- Ensembled the predicted boxes using WBF

Relationship model

- Model is LightGBM
 - Accepts input of 100 geometric features extracted from a pair of bounding boxes, *e.g. Image aspect ratios, angles, areas, box perimeters, IOU of both boxes, etc.*
 - Outputs the probability of such pair being in a relationship
 - 9 models, one for each relationship
 - Trains model with provided labeled triplet pairs (negative data are from all other pairs NOT positive)

Relationship model

- During inference, a pair of cropped bounding boxes is passed as input to the LightGBM model
- Only run inference on existing (287) labels
- Will search exclusively for all possible pairs of bounding boxes within an image, and order / filter by probabilities

Relationship model

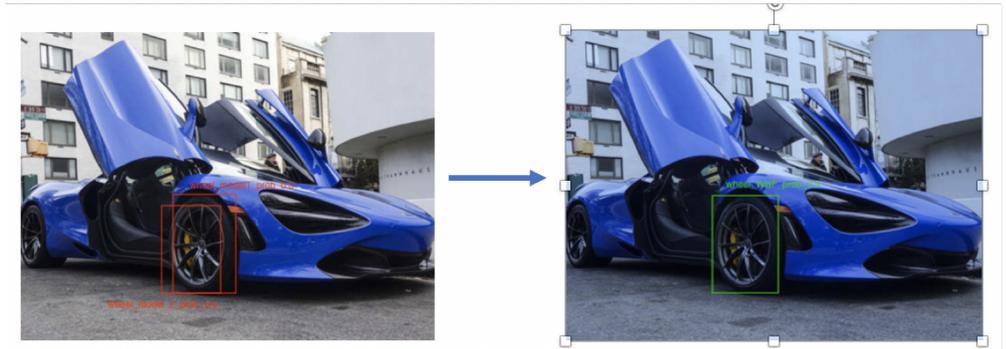
- During inference, a pair of cropped bounding boxes is passed as input to the LightGBM model
- Only run inference on existing (287) labels
- Will search exclusively for all pairs of bounding boxes within an image, and order / filtered by probabilities
- Total 3-7B possible pairs
- First filter – Confidence1 * Confidence2 -> 400M
- Second filter – Confidence1 * Confidence 2 * Prediction -> 20M

Final solution – simply concat both outputs

- Final submit is combination of both relationship model output and attribute model output

Weighted Box Fusion (WBF)

- During ensemble, there will be many boxes that have high IOU with each other
- Simply remove those boxes with low probabilities might lose useful information
- Solution -> weighted average their coordinates by box scores
- Box score = model weight * box probability
- Only apply on boxes that belong to same label AND IOU > predefined threshold
- Example:
 - $0.46 + 0.51 \sim 0.56 - (WBF) > 0.63$
 - $0.0583 + 0.0583 - (WBF) > 0.0622$
 - $0.0622 + 0.0792 + 0.0811 - (WBF) > 0.0830$



Conclusion & Future

- Sub-trains simple and effective ways of improving scores (i.e. lead to strong expert models of subsets of labels)
- More external data will definitely help the bounding boxes (i.e. object 365 as top OD teams used)
- Bounding box models performance directly affects performance of relationship model
- WBF improves performance significantly for boxes of different models (as well as mask, e.g. in Segmentation Task)

Thanks