AlignDet: Aligning Pre-training and Fine-tuning in Object Detection

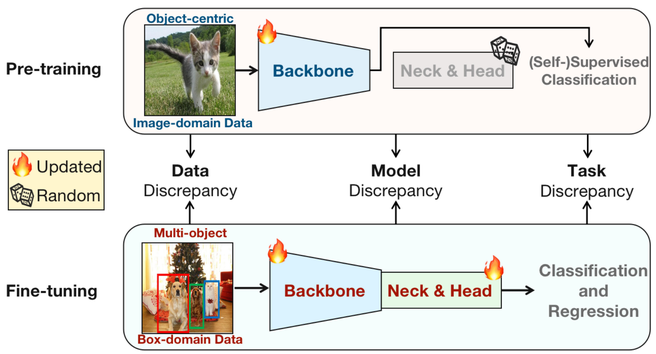
Ming Li[1]*, Jie Wu[1]*†, Xionghui Wang[1], Chen Chen[2†], Jie Qin[1], Xuefeng Xiao[1], Rui Wang[1], Min Zheng[1], Xin Pan[1]

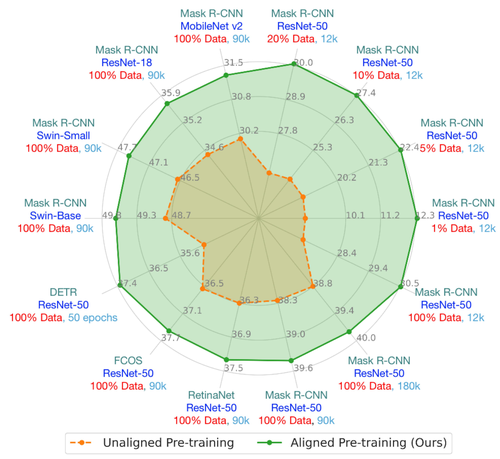
[1] TikTok, ByteDance Inc, [2] Center for Research in Computer Vision, University of Central Florida

Project arXiv Code

A General Self-supervised Pre-training for All Detection Models

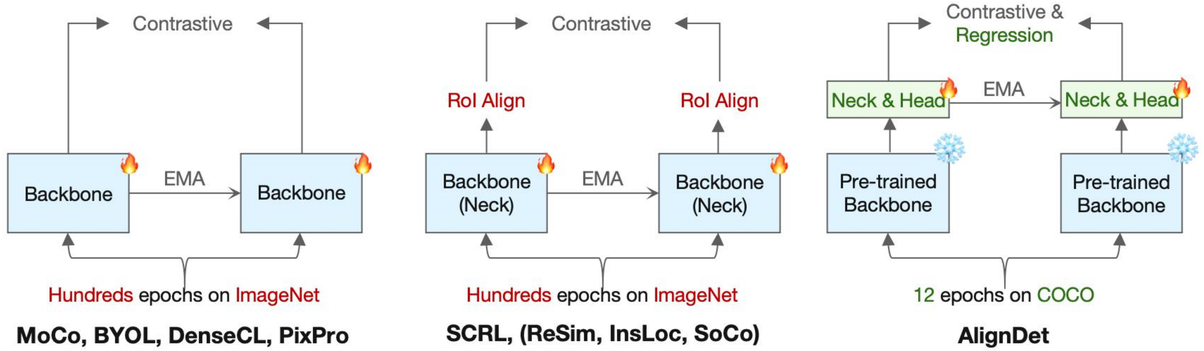
Motivation





There are **data, model, and task discrepancies** between the pre-training and fine-tuning.
Aligning these discrepancies achieves significant **improvements** across **various settings and detectors**.

Advantages

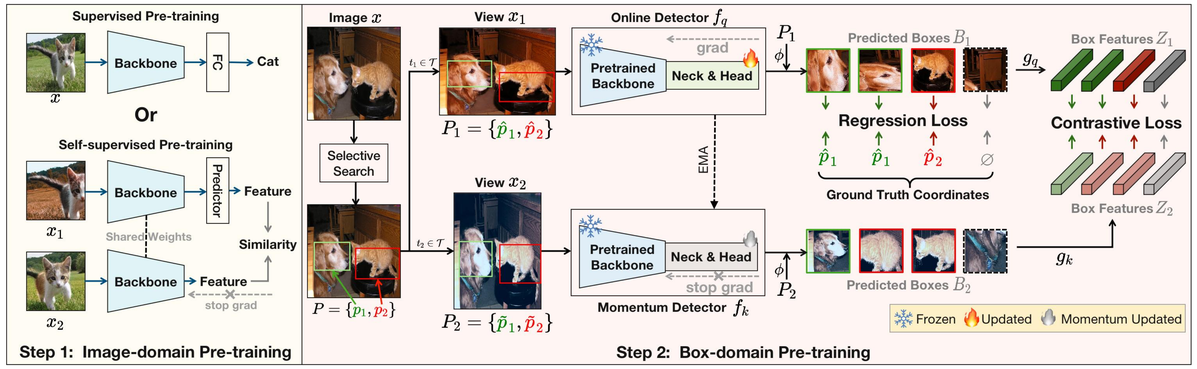


Comparison with other self-supervised pre-training methods on data, models and tasks aspects.
AlignDet achieves more **efficient, adequate and detection-oriented** pre-training.

Contributions

- ***New Insight***: We point out that existing detection algorithms are constrained by the **data**, **model**, and **task** discrepancies between pre-training and fine-tuning.
- ***Novel Method***: We propose AlignDet to align these discrepancies, which constructs **detection-oriented pre-training** by learning classification and regression knowledge.
- ***Efficiency and Pioneering***: AlignDet makes **the first attempt** to fully pre-train **all kinds of detectors** using a **completely unsupervised paradigm**, by integrating pre-trained backbones.

Pipeline



Learns Good Classification and Regression Priors

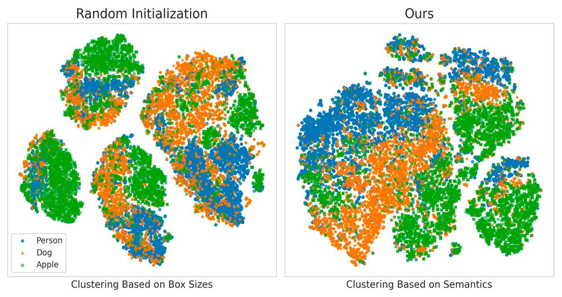


Figure 4. t-SNE visualization of ground truth annotations. AlignDet pre-training results in better class separation.

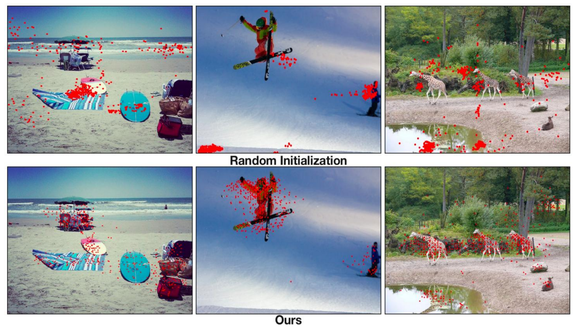


Figure 5. Visualization of predictions on COCO Val2017.