Motivations

• Ideal imitation
  • Student feature map = Teacher feature map
  • Student feature map + teacher head = correct detection result
• current state
  • The detection results are completely wrong fail if we forward the neck feature map of the student to the head of the teacher even after the feature-based and relation-based losses converge to a small value

Method

Overview

• directly forward the student’s feature map to the teacher to make the final predictions.
• Optimized the prediction of the teacher.

\[ L_{det} = L_{cls}(p, p^{\text{gt}}) + L_{reg}(t, t^{\text{gt}}) \]

\[ L_{feat} = \frac{1}{R_{\text{trans}}} \sum_{t^{\text{gt}} \notin L_{det}^{\text{gt}}} \left( \text{feat}^{t^{\text{gt}}} - \text{feat}^{t^{\text{gt}}} \right)^2 \]

\[ L = L_{det}(p, t) + \lambda_{\text{feat}} \sum_{t^{\text{gt}} \notin L_{det}^{\text{gt}}} L_{feat}(\text{feat}^{t^{\text{gt}}}, \text{feat}^{t^{\text{gt}}}) + \lambda_{\text{rel}} \sum_{t^{\text{gt}}} L_{\text{rel}}^{t^{\text{gt}}} \]

DI and FI Architecture Details

Effect of each imitation strategy in FI.

Results on MS COCO 2017 dataset.

\begin{align*}
\text{Method} & \quad \text{Imit. Strat.} & \quad \text{mAP} \\
\text{ResNeXt101(T)} & \quad - & \quad 42.7 \\
\text{MobileV2C128(S)} & \quad - & \quad 30.4 \\
\text{FCOS} & \quad \text{+Hint} & \quad 35.1_{\text{+1.1}} \\
\text{} & \quad \text{+pa} & \quad 31.8_{\text{+1.4}} \\
\text{} & \quad \text{+NonLocal} & \quad 33.3_{\text{+1.9}} \\
\text{Ours: +DI} & \quad \text{N2N} & \quad 34.7_{\text{+4.3}} \\
\text{Ours: +Dist2} & \quad \text{All2All} & \quad 36.4_{\text{+6.0}} \\
\end{align*}