Gradient Decomposition and Alignment for Incremental Object Detection

Supplementary Material

In this supplementary material, we provide mathematical proof demonstrating how GDA fosters beneficial knowledge transfer for both new and old tasks in Sec.S1. Additionally, we explain the rationale for decomposing only $\mathcal{L}_{ROI}$ in GDA while keeping $\mathcal{L}_{RPN}$ intact in Sec.S2.

S1. Mathematical Proof for GDA Does Not Prevent New Tasks Learning

As described in Sec.3.4, we propose that the latent constraint: for all $\langle \boldsymbol{g}_{proj}, \boldsymbol{g} \rangle \geq 0$ if $\lambda \in [0,1]$ (proposed in Sec.3.4), ensures that parameter updating continually drives learning forward. Consequently, the orthogonal projection strategy employed in GDA does not hinder the model's ability to learn new tasks. In this section, we provide mathematical proof to substantiate this claim. As described in Eq.(8), $\boldsymbol{g}_{proj}$ has a two-fold structure, and the proof process for both cases is outlined as follows. **(1)** If $\langle \boldsymbol{g}_{pseudo}, \boldsymbol{g} \rangle \geq 0$, $\boldsymbol{g}_{proj} = \boldsymbol{g}$, the $\langle \boldsymbol{g}_{proj}, \boldsymbol{g} \rangle \geq 0$ can be proved by:

$$\begin{aligned} \langle \boldsymbol{g}_{proj}, \boldsymbol{g} \rangle &= \langle \boldsymbol{g}, \boldsymbol{g} \rangle \\ &= 1 \text{ or } 0. \end{aligned} \tag{S1}$$

(2) If $\langle \boldsymbol{g}_{pseudo}, \boldsymbol{g} \rangle < 0$, $\boldsymbol{g}_{proj} = \boldsymbol{g} - \lambda \cdot \frac{\boldsymbol{g}_{pseudo}^\top \cdot \boldsymbol{g}}{||\boldsymbol{g}_{pseudo}||^2} \cdot \boldsymbol{g}_{pseudo}$. We first expand $\langle \boldsymbol{g}_{proj}, \boldsymbol{g} \rangle$ as:

$$\langle \boldsymbol{g}_{proj}, \boldsymbol{g} \rangle = \frac{\boldsymbol{g}_{proj}^\top \cdot \boldsymbol{g}}{||\boldsymbol{g}_{proj}|| \cdot ||\boldsymbol{g}||}. \tag{S2}$$

Since $\frac{1}{||\boldsymbol{g}_{proj}|| \cdot ||\boldsymbol{g}||} > 0$, the question is equivalent to prove $\boldsymbol{g}^\top \cdot \boldsymbol{g}_{proj} > 0$. We step forward to expand $\boldsymbol{g}^\top \cdot \boldsymbol{g}_{proj}$ as followed:

$$\begin{aligned} \boldsymbol{g}^\top \cdot \boldsymbol{g}_{proj} =& \boldsymbol{g}^\top \cdot (\boldsymbol{g} - \lambda \cdot \frac{\boldsymbol{g}_{pseudo}^\top \cdot \boldsymbol{g}}{||\boldsymbol{g}_{pseudo}||^2} \cdot \boldsymbol{g}_{pseudo}) \\ =& \boldsymbol{g}^\top \cdot \boldsymbol{g} - \lambda \cdot \boldsymbol{g}^\top \cdot \boldsymbol{g}_{pseudo} \frac{\boldsymbol{g}_{pseudo}^\top \cdot \boldsymbol{g}}{||\boldsymbol{g}_{pseudo}||^2}) \\ =& ||\boldsymbol{g}||^2 - \lambda \cdot \frac{(\boldsymbol{g}_{pseudo}^\top \cdot \boldsymbol{g})^2}{||\boldsymbol{g}_{pseudo}||^2} \\ =& \frac{(||\boldsymbol{g}|| \cdot ||\boldsymbol{g}_{pseudo}||)^2 - \lambda \cdot (\boldsymbol{g}^\top \cdot \boldsymbol{g}_{pseudo})^2}{||\boldsymbol{g}_{pseudo}||^2}. \end{aligned} \tag{S3}$$

According to Cauchy–Schwarz inequality:

$$||\boldsymbol{g}|| \cdot ||\boldsymbol{g}_{pseudo}|| \geq |\boldsymbol{g}^\top \cdot \boldsymbol{g}_{pseudo}|, \tag{S4}$$

and $||\boldsymbol{g}|| \cdot ||\boldsymbol{g}_{pseudo}|| > 0$, so:

$$(||\boldsymbol{g}|| \cdot ||\boldsymbol{g}_{pseudo}||)^2 \geq (\boldsymbol{g}^\top \cdot \boldsymbol{g}_{pseudo})^2. \tag{S5}$$

As $\lambda \in [0,1]$:

$$(||\boldsymbol{g}|| \cdot ||\boldsymbol{g}_{pseudo}||)^2 \geq \lambda \cdot (\boldsymbol{g}^\top \cdot \boldsymbol{g}_{pseudo})^2. \tag{S6}$$

At last, as $\frac{1}{||\boldsymbol{g}_{pseudo}||^2} > 0$:

$$\frac{1}{||\boldsymbol{g}_{pseudo}||^2}((||\boldsymbol{g}|| \cdot ||\boldsymbol{g}_{pseudo}||)^2 - \lambda(\boldsymbol{g}^\top \cdot \boldsymbol{g}_{pseudo})^2) \geq 0. \tag{S7}$$

In summary, $\langle \boldsymbol{g}_{proj}, \boldsymbol{g} \rangle \geq 0$ for any $\lambda \in [0,1]$, ensuring that updates based on $\boldsymbol{g}_{proj}$ continue to drive learning forward. This enhances model stability while promoting beneficial knowledge transfer for previously learned object classes.

S2. The Reasons for Only Decomposing ROI Loss in GDA

As described in Sec.3.4, to obtain loss related to old-class objects, we only decompose ROI Loss $\mathcal{L}_{ROI}$ while leaving the RPN loss $\mathcal{L}_{RPN}$ intact. The rationale for this strategy is grounded in the following reasons. **(1) Class-Agnostic Nature of RPN:** The Region Proposal Network in Faster R-CNN is class-agnostic, meaning it does not distinguish between specific object categories. Instead, RPN focuses on determining whether a region is foreground (object) or background. As a result, decomposing the RPN loss to separate old and new classes is unnecessary, as it does not contribute to class-specific knowledge. **(2) Empirical Evidence:** Experimental results, as shown in Table S1, indicate that the two strategies yield comparable performance. This demonstrates that decomposing $\mathcal{L}_{RPN}$ does not offer significant advantages while avoiding decomposition simplifies the framework. **(3) Effectiveness of GM-Pseudo:** The GM-Pseudo module effectively handles medium-confidence regions by classifying them as latent objects. These regions are excluded from being treated as negative anchors, ensuring that RPN can focus on learning new object proposals without interfering with previously learned knowledge.

Table S1. Analysis of loss decomposing in GDA. $w/o\ \mathcal{L}_{RPN}$ represents only decomposing $\mathcal{L}_{ROI}$ and $w/\ \mathcal{L}_{RPN}$ represents decomposing both $\mathcal{L}_{ROI}$ and $\mathcal{L}_{RPN}$.

Methods	5-5			10-5		
	1-5	6-15	1-20	1-10	11-20	1-20
$w/\ \mathcal{L}_{RPN}$	62.7	63.6	63.4	71.7	67.2	69.4
$w/o\ \mathcal{L}_{RPN}$	62.2	63.9	63.5	71.3	67.2	69.3

In summary, we observe the fact that the primary source of forgetting originates from the ROI Head rather than the RPN network in the incremental learning process.