Supplementary Material *for* DQS3D: Densely-matched Quantization-aware Semi-supervised 3D Detection

Huan-ang Gao[1,2] Beiwen Tian[1,2] Pengfei Li[1,2] Hao Zhao[1] Guyue Zhou[1]
[1]Institute for AI Industry Research (AIR), THU
[2]Department of Computer Science and Technology, THU
{gha20, tbw18, li-pf22}@mails.tsinghua.edu.cn {zhaohao, zhouguyue}@air.tsinghua.edu.cn

A. Proof of Equations

Lemma 1 If all elements in $\mathbf{A}$ are integers, then the following equation holds:

$$[\mathbf{A}+\mathbf{B}] = \mathbf{A} + [\mathbf{B}] \tag{1}$$

Proof: By definition. □

Lemma 2 If all elements in $\mathbf{A}$ are integers and $\theta \in \{\frac{k\pi}{2}\}_{k=0}^{3}$, then all elements in $\mathbf{A}\mathbf{R}_\theta$ are integers.

Proof: By considering the rotation matrix $\mathbf{R}_{\frac{k\pi}{2}}$ when $k = 0, 1, 2, 3$. Note that $x' = x\cos\theta + y\sin\theta$, $y' = -x\sin\theta + y\cos\theta$ and $z' = z$. When $k = 0, 1, 2, 3$, $\sin\theta$ and $\cos\theta$ produces integer values. According to the property of integer fields, x', y' and z' are also integers, which means all elements in $\mathbf{A}\mathbf{R}_\theta$ are integers. □

Proof of Equation 2 Here we prove:

$$\begin{aligned}
\tilde{\delta}_1^{\mathbf{A}'} &= \frac{\cos\theta+1}{2}\delta_1^{\mathbf{A}} + \frac{-\cos\theta+1}{2}\delta_2^{\mathbf{A}} + \frac{-\sin\theta}{2}\delta_3^{\mathbf{A}} + \frac{\sin\theta}{2}\delta_4^{\mathbf{A}},\\
\tilde{\delta}_2^{\mathbf{A}'} &= \frac{-\cos\theta+1}{2}\delta_1^{\mathbf{A}} + \frac{\cos\theta+1}{2}\delta_2^{\mathbf{A}} + \frac{\sin\theta}{2}\delta_3^{\mathbf{A}} + \frac{-\sin\theta}{2}\delta_4^{\mathbf{A}},\\
\tilde{\delta}_3^{\mathbf{A}'} &= \frac{\sin\theta}{2}\delta_1^{\mathbf{A}} + \frac{-\sin\theta}{2}\delta_2^{\mathbf{A}} + \frac{\cos\theta+1}{2}\delta_3^{\mathbf{A}} + \frac{-\cos\theta+1}{2}\delta_4^{\mathbf{A}},\\
\tilde{\delta}_4^{\mathbf{A}'} &= \frac{-\sin\theta}{2}\delta_1^{\mathbf{A}} + \frac{\sin\theta}{2}\delta_2^{\mathbf{A}} + \frac{-\cos\theta+1}{2}\delta_3^{\mathbf{A}} + \frac{\cos\theta+1}{2}\delta_4^{\mathbf{A}},\\
\tilde{\delta}_5^{\mathbf{A}'} &= \delta_5^{\mathbf{A}},\ \tilde{\delta}_6^{\mathbf{A}'} = \delta_6^{\mathbf{A}},\ \tilde{\delta}_7^{\mathbf{A}'} = \delta_7^{\mathbf{A}}\cos(2\theta),\ \tilde{\delta}_8^{\mathbf{A}'} = \delta_8^{\mathbf{A}}\cos(2\theta).
\end{aligned} \tag{2}$$

Proof: Assume the bounding box $\mathbf{y}$ is centered at $\mathbf{c} \in \mathbb{R}^{3\times 1}$ with dimension $\mathbf{d} \in \mathbb{R}^{3\times 1}$ and yaw $\phi \in \mathbb{R}$. Since the spatial translation does not affect the relative position of voxels and bounding boxes, here we can only consider the effect of random rotation around the upright-axis θ. Since we have:

$$\begin{bmatrix}\tilde{\delta}_1^{\mathbf{A}'}\\ \tilde{\delta}_2^{\mathbf{A}'}\\ \tilde{\delta}_3^{\mathbf{A}'}\\ \tilde{\delta}_4^{\mathbf{A}'}\end{bmatrix} = \begin{bmatrix}\tilde{x}-\tilde{\hat{x}}\\ \tilde{\hat{x}}-\tilde{x}\\ \tilde{y}-\tilde{\hat{y}}\\ \tilde{\hat{y}}-\tilde{y}\end{bmatrix} + \begin{bmatrix}\frac{1}{2}w\\ \frac{1}{2}w\\ \frac{1}{2}h\\ \frac{1}{2}h\end{bmatrix} = \widetilde{\begin{bmatrix}x-\hat{x}\\ \hat{x}-x\\ y-\hat{y}\\ \hat{y}-y\end{bmatrix}} + \begin{bmatrix}\frac{1}{2}w\\ \frac{1}{2}w\\ \frac{1}{2}h\\ \frac{1}{2}h\end{bmatrix} \tag{3}$$

By noting that,

$$\begin{aligned}
\widetilde{\begin{bmatrix}x-\hat{x}\\ y-\hat{y}\end{bmatrix}} &= \begin{bmatrix}\cos\theta & -\sin\theta\\ \sin\theta & \cos\theta\end{bmatrix}\begin{bmatrix}x-\hat{x}\\ y-\hat{y}\end{bmatrix}\\
&= \frac{1}{2}\begin{bmatrix}\cos\theta & -\sin\theta\\ \sin\theta & \cos\theta\end{bmatrix}\begin{bmatrix}\delta_1-\delta_2\\ \delta_3-\delta_4\end{bmatrix}\\
&= \frac{1}{2}\begin{bmatrix}\cos\theta & -\cos\theta & -\sin\theta & \sin\theta\\ \sin\theta & -\sin\theta & \cos\theta & -\cos\theta\end{bmatrix}\begin{bmatrix}\delta_1\\ \delta_2\\ \delta_3\\ \delta_4\end{bmatrix}
\end{aligned} \tag{4}$$

And that,

$$\begin{bmatrix}w\\ h\end{bmatrix} = \begin{bmatrix}\delta_1+\delta_2\\ \delta_3+\delta_4\end{bmatrix} = \begin{bmatrix}1 & 1 & 0 & 0\\ 0 & 0 & 1 & 1\end{bmatrix}\begin{bmatrix}\delta_1\\ \delta_2\\ \delta_3\\ \delta_4\end{bmatrix} \tag{5}$$

Then we have,

$$\begin{aligned}
\tilde{\delta}_1^{\mathbf{A}'} &= \frac{\cos\theta+1}{2}\delta_1^{\mathbf{A}} + \frac{-\cos\theta+1}{2}\delta_2^{\mathbf{A}} + \frac{-\sin\theta}{2}\delta_3^{\mathbf{A}} + \frac{\sin\theta}{2}\delta_4^{\mathbf{A}},\\
\tilde{\delta}_2^{\mathbf{A}'} &= \frac{-\cos\theta+1}{2}\delta_1^{\mathbf{A}} + \frac{\cos\theta+1}{2}\delta_2^{\mathbf{A}} + \frac{\sin\theta}{2}\delta_3^{\mathbf{A}} + \frac{-\sin\theta}{2}\delta_4^{\mathbf{A}},\\
\tilde{\delta}_3^{\mathbf{A}'} &= \frac{\sin\theta}{2}\delta_1^{\mathbf{A}} + \frac{-\sin\theta}{2}\delta_2^{\mathbf{A}} + \frac{\cos\theta+1}{2}\delta_3^{\mathbf{A}} + \frac{-\cos\theta+1}{2}\delta_4^{\mathbf{A}},\\
\tilde{\delta}_4^{\mathbf{A}'} &= \frac{-\sin\theta}{2}\delta_1^{\mathbf{A}} + \frac{\sin\theta}{2}\delta_2^{\mathbf{A}} + \frac{-\cos\theta+1}{2}\delta_3^{\mathbf{A}} + \frac{\cos\theta+1}{2}\delta_4^{\mathbf{A}}.
\end{aligned} \tag{6}$$

The rotation around the upright-axis does not affect z-coordinates, so it is trivial that,

$$\tilde{\delta}_5^{\mathbf{A}'} = \delta_5^{\mathbf{A}}, \tilde{\delta}_6^{\mathbf{A}'} = \delta_6^{\mathbf{A}}. \tag{7}$$

The rotation does transform the yaw angle from ϕ to $\phi - \theta$, hence we have:

$$\begin{aligned}
\tilde{\delta}_7^{\mathbf{A}'} &= \log(\frac{w}{l})\sin(2\phi-2\theta)\\
&= \log(\frac{w}{l})\left(\sin(2\phi)\cos(2\theta) - \sin(2\theta)\cos(2\phi)\right)\\
\tilde{\delta}_8^{\mathbf{A}'} &= \log(\frac{w}{l})\cos(2\phi-2\theta)\\
&= \log(\frac{w}{l})(\cos(2\phi)\cos(2\theta) + \sin(2\theta)\sin(2\phi))
\end{aligned} \tag{8}$$

By noting that when $\theta \in \{\frac{k\pi}{2}\}_{k=0}^{3}$, $\sin(2\theta) \equiv 0$. That produces,

$$\begin{aligned} \tilde{\delta}_7^{\mathbf{A}'} &= \log(\frac{w}{l})\sin(2\phi)\cos(2\theta) = \delta_7^{\mathbf{A}}\cos(2\theta) \\ \tilde{\delta}_8^{\mathbf{A}'} &= \log(\frac{w}{l})\cos(2\phi)\cos(2\theta) = \delta_8^{\mathbf{A}}\cos(2\theta) \end{aligned} \tag{9}$$

Eq. 16, Eq. 17 and Eq. 19 can be combined to form Eq. 12. □

Proof of Equation 8 Here we prove:

$$[\{\mathbf{A}\}\mathbf{R}_\theta + \{\Delta\mathbf{r}\} + \vec{\mathbf{r}'}] = \mathbf{0} \tag{10}$$

We start from Eq. 7 from the main paper:

$$[\mathbf{A}\mathbf{R}_{\theta,\Delta\mathbf{r}} + \vec{\mathbf{r}'}] = [[\mathbf{A}]\mathbf{R}_{\theta,\Delta\mathbf{r}}] \tag{11}$$

By defactoring $\mathbf{A}\mathbf{R}_{\theta,\Delta\mathbf{r}}$ into $\mathbf{A}\mathbf{R}_\theta + \Delta\mathbf{r}$, we have:

$$[\mathbf{A}\mathbf{R}_\theta + \Delta\mathbf{r} + \vec{\mathbf{r}'}] = [[\mathbf{A}]\mathbf{R}_\theta + \Delta\mathbf{r}] \tag{12}$$

Noting all elements in $[\mathbf{A}]$ are integers, hence by assuming $\theta \in \{\frac{k\pi}{2}\}_{k=0}^{3}$ and applying Lemma. 2, all elements in $[\mathbf{A}]\mathbf{R}_\theta$ are also integers. Then by Lemma. 1, we have:

$$[\mathbf{A}\mathbf{R}_\theta + \Delta\mathbf{r} + \vec{\mathbf{r}'}] = [\mathbf{A}]\mathbf{R}_\theta + [\Delta\mathbf{r}] \tag{13}$$

Leveraging the property that $\mathbf{X} = [\mathbf{X}] + \{\mathbf{X}\}$, we have:

$$[([\mathbf{A}]+\{\mathbf{A}\})\mathbf{R}_\theta+[\Delta\mathbf{r}]+\{\Delta\mathbf{r}\}+\vec{\mathbf{r}'}] = [\mathbf{A}]\mathbf{R}_\theta+[\Delta\mathbf{r}] \tag{14}$$

A simple deformation of this equation yields:

$$[[\mathbf{A}]\mathbf{R}_\theta + [\Delta\mathbf{r}] + \{\mathbf{A}\}\mathbf{R}_\theta + \{\Delta\mathbf{r}\} + \vec{\mathbf{r}'}] = [\mathbf{A}]\mathbf{R}_\theta + [\Delta\mathbf{r}] \tag{15}$$

By Lemma. 1, we move the term $[\mathbf{A}]\mathbf{R}_\theta + [\Delta\mathbf{r}]$ out of the left-hand side, and that yields:

$$[\{\mathbf{A}\}\mathbf{R}_\theta + \{\Delta\mathbf{r}\} + \vec{\mathbf{r}'}] = \mathbf{0} \tag{16}$$

That is the exact form as Eq. 8 in the original paper.□

Solution to Equation 10 Here we find the solution γ_0 of:

$$\gamma_0 = \text{argmin}_{\gamma\in[0,\mathrm{S_v}]^3}\ ||\gamma - \{\mathbf{A}\}\mathbf{R}_\theta - \{\Delta\mathbf{r}\}||_2 \tag{17}$$

We start by considering cases for unary functions. We find the solution ϕ_0 of:

$$\phi_0 = \text{argmin}_{\phi\in[\mathrm{a,b}]}\ ||\phi - \mathrm{M}||_2 \tag{18}$$

The solution is straight-forward. It denotes the closest value in $[a, b]$ to a fixed value M. We represent the solution to this problem as:

$$\mathrm{clamp(M, a, b)} = \begin{cases} a, & M < a, \\ M, & a \le M < b, \\ b, & M \ge b. \end{cases} \tag{19}$$

Since in the target function of this problem, the three axes are uncorrelated, we can break this problem to a problem set of three problems each equivalent to Eq. 10. We can extend the clamping function to a vector version, namely for any $0 \le i < \mathrm{len}(\mathbf{M})$:

$$\mathrm{clamp}(\mathbf{M}, \mathbf{a}, \mathbf{b})_\mathrm{i} = \mathrm{clamp}(\mathbf{M}_\mathrm{i}, \mathbf{a}_\mathrm{i}, \mathbf{b}_\mathrm{i}) \tag{20}$$

Then the closed-form solution of γ_0 can be formulated as:

$$\gamma_0 = \mathrm{clamp}(\mathbf{M} = \{\mathbf{A}\}\mathbf{R}_\theta + \{\Delta\mathbf{r}\}, (0, 0, 0), (\mathrm{S_v}, \mathrm{S_v}, \mathrm{S_v})) \tag{21}$$

That is the solution to the original problem. □

B. Hyperparameter Study

τ_{center} and τ_{cls}. We conducted a hyperparameter study (Fig. 1) on τ_{center} and τ_{cls}. These two hyperparameters are utilized to filter the initially matched set and provide matching pairs that offer less noisy supervision. Finding the optimal values involves a trade-off, as setting the values too low introduces noisy supervision, while setting them too high reduces the number of matched pairs.

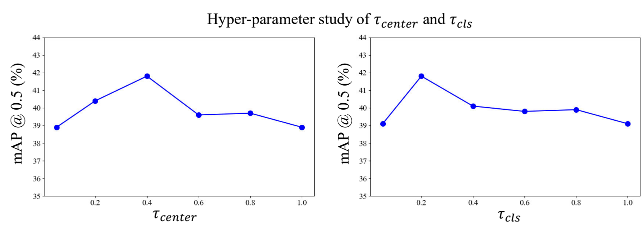

Figure 1: Hyper-parameter Study on τ_{center} and τ_{cls}.

Backbone (Semi-supervised Setting)	mAP@0.25	mAP@0.50
FCAF3D (baseline)	58.2	42.1
FCAF3D (+ Sparse Proposal Matching)	62.0	44.2
FCAF3D (+ Dense Matching, *ours DQS3D*)	**64.3**	**48.5**
TR3D (baseline)	62.5	46.8
TR3D (+ Dense Matching)	**65.4**	**49.9**

Table 1: Comparison of *Dense Matching* and *Proposal Matching* Strategies with *Different Backbones* on ScanNet Dataset (20% Labeled). *Proposal matching* involves filtering teacher proposals and matching them with the nearest-center student predictions, while dense matching establishes matching based on spatially-aligned voxel anchors and then applies filtering. In dense matching, the proposed **Q**uantization **E**rror **C**orrection module is enabled.

Different Backbones. We conducted experiments (Tab. 1) that show the superiority of dense matching over proposal matching. We argue that the success is originate from addressing issues like *no supervision* and *multiple supervision* problems, which we also qualitatively illustrate in Fig. 6. Note that dense matching is applicable only to recent SOTA voxel-based detectors, not common two-stage proposal-based detectors based on Transformer or heatmaps. Hence we used TR3D (Rukhovich et al.), with the hyperparameters reported in our manuscript without further tuning. Remarkably, we observed an improvement of +3.1% on $\mathrm{mAP@0.50}$.

C. Further Discussion

Computational Complexity Analysis. We utilized the NVIDIA GeForce RTX 2080Ti. Training employed 4 GPUs (2 labeled and 2 unlabeled scenes per GPU card, occupying approximately 7.5GB per GPU) and took around 7 hours to converge. In terms of inference speed, our system achieves 10.3 scenes per second on a single 2080Ti.

Limitation Analysis. The trade-off between memory and voxel size hampers our 3D detectors' performance in outdoor scenes, which is a common limitation in the family of sparse convolutional detectors.