

Supplementary Material: 6DoF Object Pose Estimation via Differentiable Proxy Voting Regularizer

Xin Yu¹

xin.yu@uts.edu.au

Zheyu Zhuang²

zheyu.zhuang@anu.edu.au

Piotr Koniusz^{2,3}

piotr.koniusz@data61.csiro.au

Hongdong Li²

hongdong.li@anu.edu.au

¹ University of Technology Sydney² Australian National University³ Data61, CSIRO, Australia

1 Derivation of Eqn. (2) from the main submission

In Eqn. (2), the first line indicates that we measure the distance between the keypoint \mathbf{k} and its corresponding foot of the perpendicular $\mathbf{f}_k(\mathbf{p})$. Since there is a close form solution for the distance from a point to a line [1], the distance d is calculated by,

$$d = \frac{|Ax^* + By^* + C|}{\sqrt{A^2 + B^2}}, \quad (1)$$

where a line is defined as $Ax + By + C = 0$ and a point is expressed as (x^*, y^*) .

We first obtain the line equation expressed by the direction $\mathbf{v}_k(\mathbf{p}) = (v_k^x, v_k^y)$ and a point $\mathbf{p} = (p^x, p^y)$ as follows,

$$\frac{x - p^x}{y - p^y} = \frac{v_k^x}{v_k^y}.$$

Then, we have:

$$v_k^y x - v_k^x y + v_k^x p^y - v_k^y p^x = 0. \quad (2)$$

Finally, substituting the keypoint coordinates $\mathbf{k} = (k^x, k^y)$ and Eqn. (2) into Eqn. (1), we achieve,

$$d = \frac{|v_k^y k^x - v_k^x k^y + v_k^x p^y - v_k^y p^x|}{\sqrt{(v_k^x)^2 + (v_k^y)^2}}. \quad (3)$$

The equations above lead to Eqn. (2) in the paper.

2 More qualitative results

In this supplementary material, we provide more visual results of our method.



Figure 1: Visualization of qualitative results on the **LINEMOD** dataset. Green 3D bounding-boxes represent the ground-truth poses and the blue ones indicate our predictions.

References

- [1] https://en.wikipedia.org/wiki/Distance_from_a_point_to_a_line



Figure 2: Visualization of qualitative results on the LINEMOD dataset. Green 3D bounding-boxes represent the ground-truth poses and the blue ones indicate our predictions.



Figure 3: Visualization of qualitative results on the **LINEMOD** dataset. Green 3D bounding-boxes represent the ground-truth poses and the blue ones indicate our predictions.



Figure 4: Visualization of qualitative results on the **LINEMOD** dataset. Green 3D bounding-boxes represent the ground-truth poses and the blue ones indicate our predictions.



Figure 5: Visualization of qualitative results on the **Occlusion LINEMOD** dataset. Green 3D bounding-boxes represent the ground-truth poses and the blue ones indicate our predictions.

