Improving Robotic Grasping on Monocular Images Via Multi-Task Learning and Positional Loss

William Prew[1,2], Toby Breckon[1], Magnus Bordewich[1], and Ulrik Beierholm[2]

Department of Computer Science[1] & Psychology[2]
Durham University

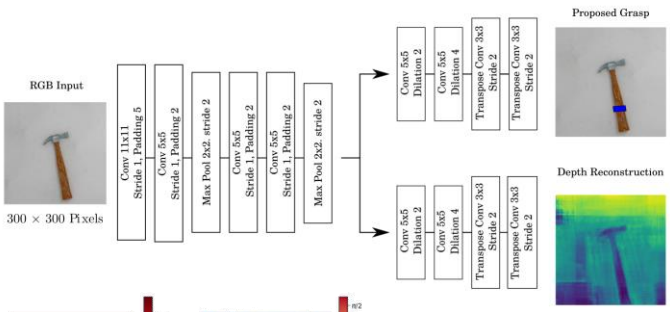
Introduction

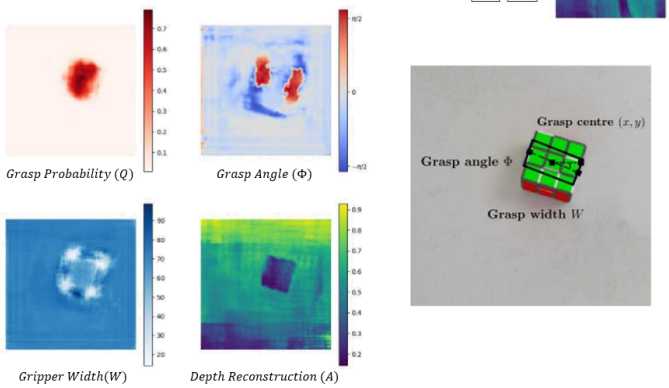
Robotic grasping, like many data-driven challenges, has recently applied **machine learning** to generate accurate grasp plans.

Learning similar concurrent tasks during training has been shown to **improve performance** on a primary task [1].

Multi-Task Grasping Convolutional Neural Network (MTG-CNN)

Shares backbone architecture with the Generative Grasping Convolutional Neural Network (GG-CNN) [2]



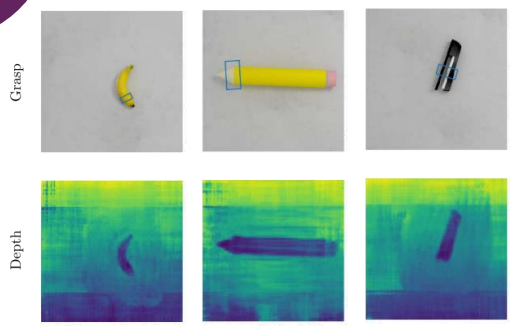


Positional Loss

We also introduce a **new loss function** which we term the **positional loss,** referenced by subscript $\boldsymbol{p}$.

$$L_p = \frac{1}{N}\sum(Q_\theta - Q_{GT})^2 + \frac{1}{N}\sum(Q_{GT}(\Phi_\theta^{sin} - \Phi_{GT}^{sin}))^2 + \frac{1}{N}\sum(Q_{GT}(\Phi_\theta^{cos} - \Phi_{GT}^{cos}))^2 + \frac{1}{N}\sum(Q_{GT}(W_\theta - W_{GT}))^2$$

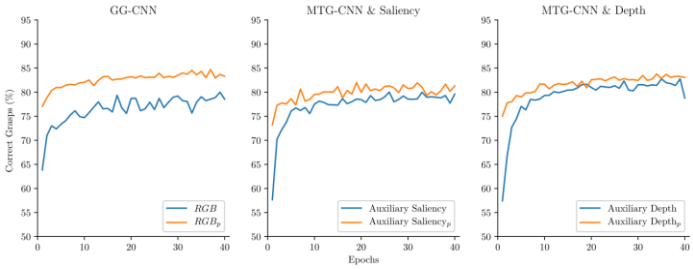
The typical **MSE loss** of the grasp**,** represented by **grasp position** (Q), **angle** (ϕ^{sin}/ϕ^{cos}) and **gripper width** (W), is scaled by per pixel grasp position ground truth (Q_{GT}).



Example grasps and disparity maps from MTG-CNN

Results

All networks are trained and evaluated on the **Jacquard grasping dataset** [3].



The average performance of each model during training with and without the positional loss function included during training.

TABLE I
THE MEAN PERCENTAGE OF CORRECT GRASPS FOR EACH NETWORK EVALUATED ON THE JACQUARD GRASPING DATASET.

Model	Auxiliary Task	Successful Grasps
GG-CNN		$72.04\% \pm 3.44$
GG-CNN$_p$		$78.92\% \pm 0.97$
MTG-CNN	Saliency	$74.93\% \pm 1.86$
MTG-CNN$_p$	Saliency	$76.23\% \pm 2.75$
MTG-CNN	Depth	$78.14\% \pm 0.65$
MTG-CNN$_p$	**Depth**	$\mathbf{79.12\% \pm 1.40}$

Conclusions

- Learning **associated concurrent tasks improves grasping performance** on the Jacquard dataset.
- The **positional loss function** also slightly improves grasp performance but also **encourages faster learning in the earlier epochs of training**.

References
[1] Ruder, S. (2017). An Overview of Multi-Task Learning in Deep Neural Networks. *ArXiv*. Retrieved from http://arxiv.org/abs/1706.05098
[2] Morrison, D., Corke, P., & Leitner, J. (2020). Learning robust, real-time, reactive robotic grasping. *International Journal of Robotics Research*, *39*(2–3),
[3] Depierre, A., Dellandrea, E., & Chen, L. (2018). Jacquard: A Large Scale Dataset for Robotic Grasp Detection. In *IEEE International Conference on Intelligent Robots and Systems* (pp. 3511–3516).