Supplementary Material for: What Matters For Meta-Learning Vision Regression Tasks?

Ning Gao[1,2] Hanna Ziesche[1] Ngo Anh Vien[1] Michael Volpp[2] Gerhard Neumann[2]
[1]Bosch Center for Artificial Intelligence [2]Autonomous Learning Robots, KIT
{ning.gao, hanna.ziesche}@de.bosch.com anhvien.ngo@bosch.com
{michael.volpp, gerhard.neumann}@kit.edu

A. Appendix

A.1. Functional Contrastive Learning on CNPs

τ	1.0	0.5	0.2	0.07	0.007
IC	8.5550	8.9810	8.8551	**7.8196**	8.1409
CC	10.4660	10.5135	10.5604	**8.8420**	9.3846

Table 1. Results of the evaluation on ShapeNet1D using different temperature values in FCL.

τ	1.0	0.5	0.2	0.07	0.007
IC	0.1564	0.174	0.1962,	0.1441	**0.1401**
CC	0.1594	0.1758	0.2089	0.1390	**0.1332**

Table 2. Results of the evaluation on ShapeNet2D using different temperature values in FCL.

Methods	Max	Max_{FCL}
ARI ↑	0.21	0.20
MI ↑	1.13	1.03
SS ↑	0.31	0.15
CHI ↑	118.73	18.90
DBI ↓	1.00	1.65

Table 3. Analysis of latent task representation on Distractor between Max and Max_{FCL} using various clustering metrics.

A grid search on hyperparameter τ is very expensive especially on vision tasks. Therefore, we search only on a discrete set $\{0.007, 0.7, 0.2, 0.5, 1.0\}$ and find that $\tau = 0.07$ shows the best performance on ShapeNet1D and $\tau = 0.007$ on ShapeNet2D. The results are shown in Tab. 1 and Tab. 2.

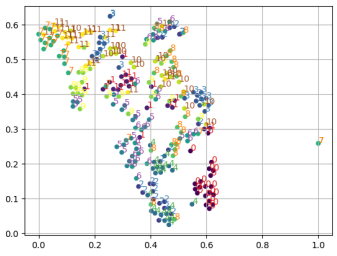
For the Distractor task, we visualize the task representation obtained for novel objects in Fig. 1 where each color or number denotes one category and each point denotes the representation of each novel object. $\{10, 11\}$ are the novel categories $\{$*sofa, watercraft*$\}$. Note that each object is considered as a single task and all tasks are learned in a category-agnostic manner. This figure indicates that Max_{FCL} can better shrink the distance between similar objects and repel the different ones implicitly. For instance, without a contrastive loss there is one outlier in Fig. 1a that is far away in representation space from the other objects. In particular, some samples are not well clustered based on categories, which is due to the high object variations within the same category.

Furthermore, we investigate the influence of FCL on the predicted task representations over all 12 categories using five clustering metrics, namely Adjusted Rand Index (ARI), Mutual Information (MI), Silhouette Score (SS), Calinski-Harabasz Index (CHI) and Davies-Bouldin Index (DBI). Results are shown in Tab. 3. FCL leads to a more dispersed latent distribution compared to the original CNP, which reduces the vacancy in the latent space and thus improve the generalization ability to unseen tasks.

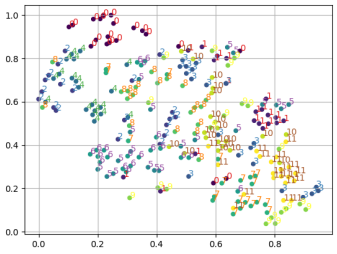
A.2. Training Details

For all tasks, we use 500k training iterations for CNPs and 70k for MAML. Furthermore, the best model on the intra- and cross-category dataset is saved during training. This leads to better models than early stopping with manually defined intervals. All experiments are conducted on a single NVIDIA V100-32GB GPU. Distractor and ShapeNet2D need around $3-5$ days for training, depending on different choices of augmentations, Pascal1D needs 8 hours and ShapeNet1D around 12 hours.

Additional Results. We have evaluated MMAML [2], a conditional variant of MAML, on ShapeNet1D based on reviewer's recommendation in Tab. 4. The results is worse than MAML, indicating that the designed task-aware modulation in MMAML doesn't benefit our tasks.



(a) Max



(b) Max_{FCL}

Figure 1. Visualization of latent variables on (a) max aggregation (b) max aggregation + functional contrastive learning (Max_{FCL}).

MMAML	No Aug	DA	TA	DA+TA
IC	19.6900	26.3624	19.0705	27.4973
CC	20.6123	26.4090	19.4285	27.3120

Table 4. Performance of MMAML [2] on ShapeNet1D.

A.3. Task Augmentation

The angular orientation of Pascal1D is normalized to $[0, 10]$ whereas ShapeNet1D uses radians with range $[0, 2\pi]$. For ShapeNet2D, the azimuth angles are restricted to the range $[0°, 180°]$ in order to reduce the effect of symmetric ambiguity while elevations are restricted to $[0°, 30°]$. we add random noise to both azimuth and elevation angles and then convert the rotation to quaternions for training.

A.4. Data Augmentation

Affine scales images between $80\% - 120\%$ of their size along x and y axis and translate the images between $-10\% - 10\%$ relative to the image height and width, and fills random value for the newly created pixels. *Dropout* either drops random 1%-10% of all pixels or random image patches with $2\% - 25\%$ of the original image size. *CropAndPad* pads each side of the images less than 5% of the image size using random value or the closest edge value. For ShapeNet2D, we furthermore add *GammaContrast* with a range $[0.5, 2.0]$, *AddToBrightness* with a range $[-30, 30]$ and *AverageBlur* using a window of $k \times k$ neighbouring pixels where $k \in [1, 3]$. We use the open-source package [1] for all data augmentations.

A.5. Meta Regularization

Yin et al. [3] employ regularization on weights, the loss function is defined as:

$$\mathcal{L} = \mathcal{L}_O + \beta D_{\text{KL}}(q(\theta; \theta_\mu, \theta_\sigma)||r(\theta)) \quad (1)$$

where L_O denotes the original loss function defined individually in Distractor and pose estimation. meta-parameters θ denote the parameters which are not used to adapt to the task training data. Function $r(\theta)$ is a variational approximation to the marginal which is set to $\mathcal{N}(\theta; 0, I)$ in Yin et al. [3]. We follow the same setup in our experiments.

A.6. Examples of Inference Results

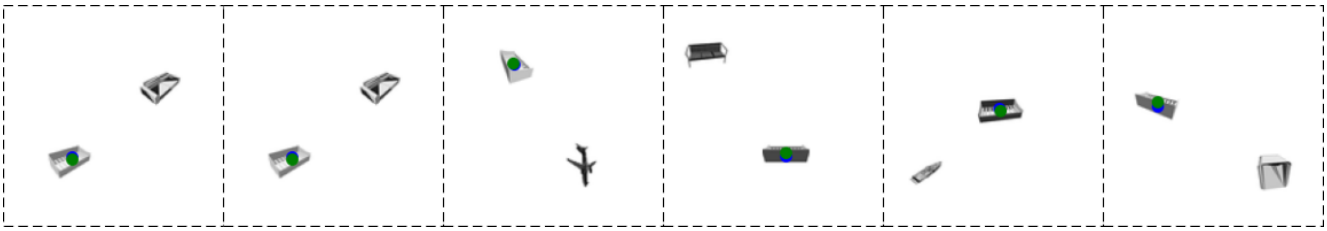
We visualize examples of evaluation on novel cateogories in Fig. 2 for Distractor, Fig. 3 for ShapeNet1D and Fig. 4 for ShapeNet2D.

References

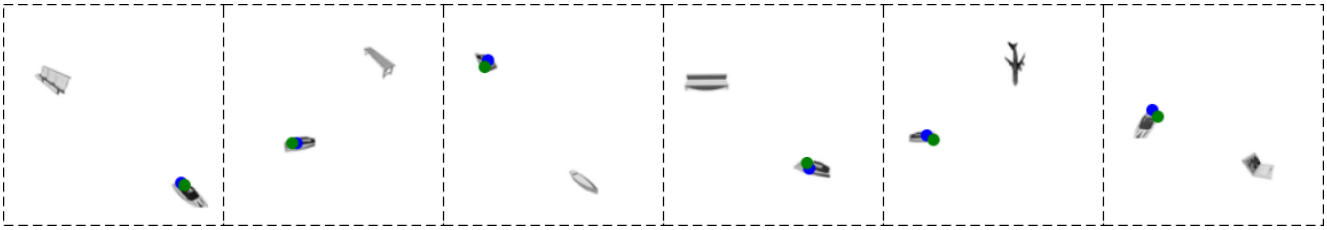
[1] Alexander B. Jung, Kentaro Wada, Jon Crall, Satoshi Tanaka, Jake Graving, Christoph Reinders, Sarthak Yadav, Joy Banerjee, Gábor Vecsei, Adam Kraft, Zheng Rui, Jirka Borovec, Christian Vallentin, Semen Zhydenko, Kilian Pfeiffer, Ben Cook, Ismael Fernández, François-Michel De Rainville, Chi-Hung Weng, Abner Ayala-Acevedo, Raphael Meudec, Matias Laporte, et al. imgaug. https://github.com/aleju/imgaug, 2020. Online; accessed 01-Feb-2020.

[2] Risto Vuorio, Shao-Hua Sun, Hexiang Hu, and Joseph J Lim. Multimodal model-agnostic meta-learning via task-aware modulation. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019.

[3] Mingzhang Yin, George Tucker, Mingyuan Zhou, Sergey Levine, and Chelsea Finn. Meta-learning without memorization. In *International Conference on Learning Representations*, 2020.

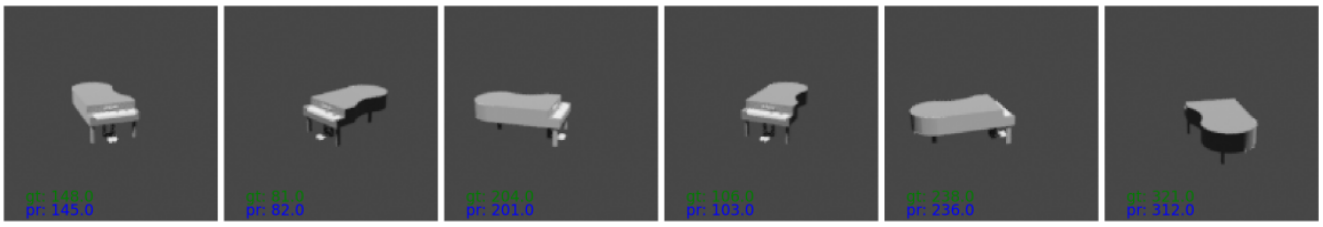


(a) Distractor sofa

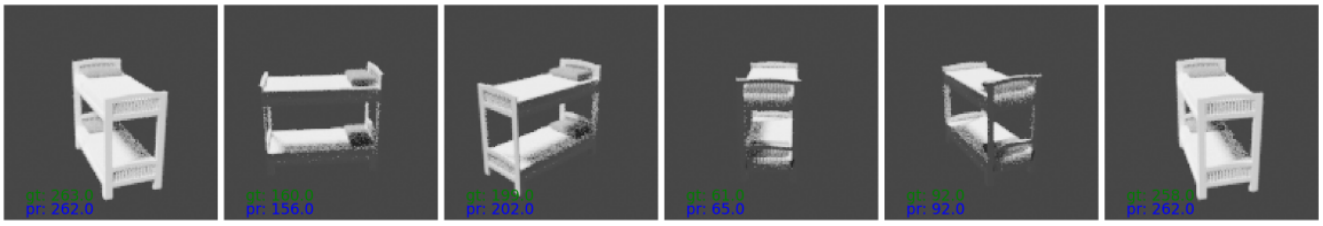


(b) Distractor watercraft

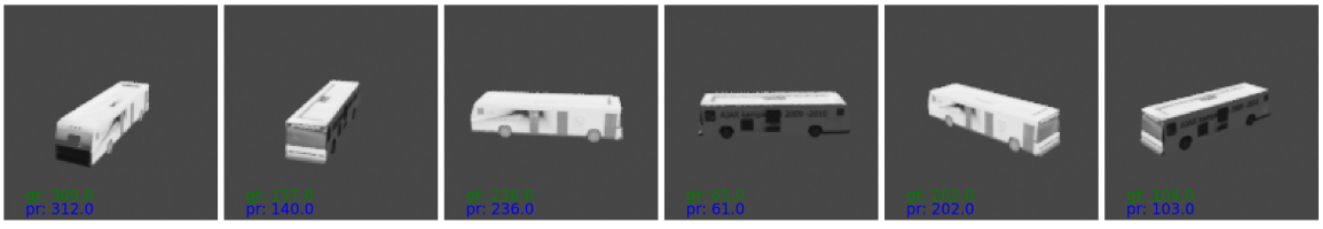
Figure 2. Examples of Distractor on novel categories (sofa and watercraft) where green dots are ground-truth and blue dots are predicted positions.



(a) ShapeNet1D piano



(b) ShapeNet1D bed



(c) ShapeNet1D bus

Figure 3. Examples of ShapeNet1D on novel categories (piano, bed, bus).

Figure 4. Examples of ShapeNet2D on novel categories (piano, bed, bus). Predictions are converted to (azimuth, elevation) angles.