MetaAug: Meta-Data Augmentation for Post-Training Quantization Supplementary Materials

Cuong Pham[1], Hoang Anh Dung[1], Cuong C. Nguyen[2], Trung Le[1], Dinh Phung[1,3], Gustavo Carneiro[2], and Thanh-Toan Do[1]

[1] Department of Data Science and AI, Monash University, Australia
[2] Academic and research departments, University of Surrey, England
[3] VinAI, Vietnam

1 Hyper-parameter settings

***Hyper-parameters* λ_1, λ_2, λ_3.** Regarding hyper-parameters λ_1, λ_2, and λ_3 in Eq. (15) in the main paper, these parameters control the impacts of validation loss, margin loss, and preservation loss on the overall loss for learning the transformation network T. We present ablation studies on the choice of hyper-parameters λ_1, λ_2, and λ_3 on the ImageNet dataset. For all experiments in this supplementary material, $\mathcal{L}_{DP}$ is used in the $\mathcal{L}_T$ (Eq. (15) in the main paper).
For ablation studies for parameter λ_1, we vary the value of λ_1 from 1 to 10 and fix the value of $\lambda_2 = 0$, and $\lambda_3 = 3 \times 10^4$. The results are shown in Table A.1. The results show that λ_1's range from 5 to 10 often leads to better performance for the 2/2 and 2/4 settings, and the proposed method does not show high sensitivity to the choice of λ_1.

Table A.1: Ablation study for hyper-parameter λ_1 of validation loss in Eq. (15). The results are on the ImageNet dataset with 2/2 and 2/4 settings.

λ_1	1	2	3	5	8	10
2/2	54.06	54.03	53.99	**54.09**	54.01	54.05
2/4	65.88	65.94	65.91	65.96	65.93	**66.03**

For ablation studies for parameter λ_2, we vary the value of λ_2 from 0.1 to 1, and fix the value of $\lambda_1 = 5$, and $\lambda_3 = 3 \times 10^4$. The ϵ in Eq. (14) is set to 0.1. The results are shown in Table A.2. The results indicate that λ_2's range from 0.2 to 0.5 yields better performance.
For ablation studies for parameter λ_3, we vary the value of λ_3 from 10^4 to 10^5, and fix the value of $\lambda_1 = 5$, and $\lambda_2 = 0.5$. The ϵ in Eq. (14) is set to 0.1. The results are shown in Table A.3. The results show that the λ_3's range from 2×10^4 to 5×10^4 often leads to higher performance, while the performance may not be sensitive to the choice of λ_3.

Table A.2: Ablation study for hyper-parameter λ_2 of margin loss in Eq. (15). The results are on the ImageNet dataset with 2/2 and 2/4 settings.

λ_2	0.1	0.2	0.3	0.5	0.8	1.0
2/2	54.02	53.90	54.17	**54.22**	54.15	54.19
2/4	65.90	**66.05**	65.80	66.01	65.83	65.97

Table A.3: Ablation study for hyper-parameter λ_3 of distribution preservation loss in Eq. (15). The results are on the ImageNet dataset with 2/2 and 2/4 settings.

λ_3	1×10^4	2×10^4	3×10^4	5×10^4	8×10^4	10×10^4
2/2	54.10	54.04	**54.22**	54.12	54.08	54.15
2/4	65.87	**66.05**	66.01	66.02	65.81	65.74

The sensitivity of hyper-parameter ϵ in Eq. (14). We conduct ablation study for the sensitivity of hyper-parameter ϵ. We vary the value of ϵ from 0.1 to 2 and fix the value of $\lambda_1 = 5$, $\lambda_2 = 0.5$, and $\lambda_3 = 3 \times 10^4$. The results are presented in Table A.4. The results show that the best value of ϵ is 0.3 for the 2/2 setting and 0.1 for the 2/4 setting. Setting ϵ higher (e.g., $\epsilon = 2$) results in modified images that could not retain the intrinsic information from the original images.

Table A.4: Ablation study for hyper-parameter ϵ of margin loss in Eq. (15). The results are on the ImageNet dataset with 2/2 and 2/4 settings.

ϵ	0.1	0.2	0.3	0.5	0.8	1	2
2/2	54.22	54.03	**54.44**	53.86	54.02	54.10	53.90
2/4	**66.01**	66.95	65.91	65.87	65.86	65.75	65.61

2 Additional comparisons with automated data augmentation

In addition to traditional augmentation techniques (e.g. Random Flip, Rotation, Brightness) and advanced augmentation methods (e.g. MixUp, CutMix) that have been presented in the main paper, we also compare the results of MetaAug with automated data augmentation approaches including RandAugment [1], and TrivialAugment [3]. These augmentations are combinations of multiple transforms, either geometric or photometric, or both. Following [1,3], we adopt the 14 different transformations: *identity, autocontrast, equalize, posterize, rotate, solarize, shear-x, shear-y, translate-x, translate-y, color, contrast, brightness, and sharpness*. Among those transformations, the photometric transformations include: *autocontrast, equalize, posterize, solarize, color, contrast, brightness, and*

Table A.5: Comparative Top-1 classification accuracy (%) on the ImageNet dataset with the 2/2 setting with ResNet-18 between our proposed method and automated data augmentation.

Augmentation	**Bit-width** (W/A)	**ResNet-18** (FP: 71.01)
Genie-M (no augmentation) [2]	2/2	53.71
TrivialAugment [3]		53.86
RandAugment [1]		53.55
MetaAug (Ours)		**54.22**
MetaAug (Ours) + TrivialAugment		54.06
MetaAug (Ours) + RandAugment		53.92

sharpness. Meanwhile, the geometric transformations include: *rotate, shear-x, shear-y, translate-x, and translate-y.*

Automated data augmentation. We first compare the proposed MetaAug with automated data augmentation approaches using 14 transformations that include both photometric and geometric transformations. The results presented in Table A.5 show that TrivalAugment and RandAugment seem not to impact the original Genie-M [2], and the performance is even decreased with RandAugment. Additionally, the combination of images produced by those methods and images produced by our transformation network also leads to performance decreases.

Table A.6: Comparative Top-1 classification accuracy (%) on the ImageNet dataset with the 2/2 setting with ResNet-18 between our proposed method and automated photometric data augmentation.

Augmentation	**Bit-width** (W/A)	**ResNet-18** (FP: 71.01)
Genie-M (no augmentation) [2]	2/2	53.71
TrivialAugment (photometric) [3]		53.53
RandAugment (photometric) [1]		53.46
MetaAug (Ours)		**54.22**
MetaAug (Ours) + TrivialAugment (photometric)		53.89
MetaAug (Ours) + RandAugment (photometric)		53.80

Automated photometric data augmentation. Table A.6 shows the results when automated data augmentation only contains the photometric transformations. The results indicate that the combination of images produced by automated photometric data augmentation and images produced by our transformation network results in performance decreases. In addition, automated photometric data augmentation methods result in performance decreases of 0.18%

Table A.7: Comparative Top-1 classification accuracy (%) on ImageNet dataset with the 2/2 setting with ResNet-18 between our proposed method and automated geometric data augmentation.

Augmentation	**Bit-width** (W/A)	**ResNet-18** (FP: 71.01)
Genie-M (no augmentation) [2]	2/2	53.71
TrivialAugment (Geometric) [3]		54.04
RandAugment (Geometric) [1]		54.06
MetaAug (Ours)		54.22
MetaAug (Ours) + TrivialAugment (Geometric)		**54.52**
MetaAug (Ours) + RandAugment (Geometric)		54.41

and 0.25% over baseline Genie-M [2] for the TrivialAugment [3] and RandAugment [1] settings, respectively. This indicates that simple photometric augmentation could potentially reduce the performance of PTQ.

Automated geometric data augmentation. Table A.7 shows the result when automated data augmentation contains only the combination of the geometric transformations. The results show that these augmentation techniques can enhance the performance of PTQ. Specifically, using automated geometric data augmentation achieves improvements over the baseline Genie-M [2] by 0.33% and 0.35% for TrivialAugment and RandAugment, respectively, in the 2/2 setting. Combining the images produced by our MetaAug with images produced by automated geometric augmentation, as shown in Table A.7, leads to a significant enhancement in PTQ performance, achieving the highest results in this table. The improvements over the baseline Genie-M (no augmentation) [2] are 0.81% and 0.70% for TrivialAugment and RandAugment, respectively, in the 2/2 setting. Meanwhile, the improvements over MetaAug alone are 0.3% and 0.19% for TrivialAugment and RandAument, respectively. This indicates that our approach MetaAug and automated geometric data augmentation can complement each other when used together.

Table A.8: The comparative performance of PTQ with various calibration data sizes on ResNet-18 in the 2/2 setting.

Num. Images	**32**	**64**	**128**	**256**	**512**
Genie-M [2]	16.17	33.13	42.29	48.37	51.50
MetaAug (Ours)	23.79	37.71	44.25	49.07	52.32

3 Efficacy for various calibration data sizes

We validate the effectiveness of our proposed method using various calibration data sizes, from 32 to 512 images. Table A.8 shows that our method consistently

Fig. A.1: Visualization of the original calibration images (the first and third rows) and the corresponding modified images (the second and fourth rows) produced by the transformation network.

outperforms Genie-M [2], and the larger improvements are achieved with smaller calibration data sizes, e.g., the improvements are 7.62% and 4.58% with 32 and 64 calibration images, respectively. This demonstrates the effectiveness of our proposed method, especially in challenging conditions with limited data.

4 More visualization as Fig. 1 in the main paper

Fig. A.1 shows the visualization of the original images and the modified images using the proposed MetaAug. The results show that the modified images change the appearance of the original images while still preserving the semantic information of the original images.

References

1. Cubuk, E.D., Zoph, B., Shlens, J., Le, Q.V.: Randaugment: Practical automated data augmentation with a reduced search space. In: CVPR. pp. 702–703 (2020) 2, 3, 4
2. Jeon, Y., Lee, C., Kim, H.y.: Genie: Show me the data for quantization. In: CVPR (2023) 3, 4, 5
3. Müller, S.G., Hutter, F.: Trivialaugment: Tuning-free yet state-of-the-art data augmentation. In: ICCV (2021) 2, 3, 4