Supplementary Material for: Improving Diversity with Adversarially Learned Transformations for Domain Generalization

Tejas Gokhale* § Rushil Anirudh # Jayaraman J. Thiagarajan # Bhavya Kailkhura #
Chitta Baral § Yezhou Yang §
[1] Arizona State University [2] Lawrence Livermore National Laboratory
{tgokhale, chitta, yz.yang}@asu.edu {anirudh1, jjayaram, kailkhura1}@asu.edu

This document contains training settings, additional results and visualizations to supplement the main paper. We also discuss the limitations of ALT and scope for future work in this direction. Code to reproduce experiments has been released publicly: https://github.com/tejas-gokhale/ALT.

1. Training Settings

Table 1 shows the training settings and hyperparameters used for experiments on each benchmark. See Algorithm 1 in the main paper for context and relevant equations.

2. Detailed Results

We provide detailed results including standard deviation values for our models on the PACS and Office-Home benchmarks, for each source domain. We compare these with RandConv [5], AugMix [3], and a combination of RandConv and AugMix which utilizes AugMix as one of the two augmentations in the consistency constraint of RandConv. Results of the PACS experiments are shown in Tables 3, 4, 5, 6, when using P, A, C, and S as the source datasets, respectively. Results of the Office-Home experiments are show in Tables 7, 8, 9, 10, when using R, A, C, and P as the source datasets, respectively.

3. Visualizations

In this section, we provide additional visualizations and qualitative examples for augmented images generated by ALT, for Digits (Figure 4), PACS (Figures 5, 6, 7, 8), and Office-Home PACS (Figures 9, 10, 11, 12). In each figure, the first row shows input images x, the second row shows the outputs of the diversity module $r(\mathrm{x})$, and the third row shows the outputs of the adversary network $g(\mathrm{x})$.

*Work done during internship at LLNL

In Figure 2 and 3 we show an illustration of the diversity introduced by ALT in comparison to the source distribution, the target (OOD) distribution, and the distribution of RandConv augmentations, for the PACS and OfficeHome benchmarks respectively. The diversity introduced by ALT is much larger and wide-spread than data augmentation techniques such as RandConv.

4. Limitations and Future Directions

In this paper we have explored the effectiveness of ALT on three standard domain generalization benchmarks. For fair comparison, for each baseline model, we use the same model architecture and training settings as the backbone – for instance, ERM, RandConv, AugMix and ALT are all trained with the same hyperparameters as shown in Table 1. For significance of results, we have repeated each experiment (including those in analyses) for 5 different seeds and have reported mean and standard deviation.

4.1. Complexity of Adversary Network.

One limitation (and therefore scope for future work) is that we have considered one family of architecture for our adversary network g – fully convolutional image-to-image translation networks. We conduct additional analysis to understand how this choice affects generalization performance, and compare performance when using between 2 and 6 convolutional layers. We reuse all other training settings from our benchmark model $ALT_{RandConv}$ on both Digits and PACS. Results are shown in Table 2.

For PACS and OfficHome, we observe that all ALT models compared are better than previous baselines including AugMix and RandConv. For Digits, we observe that performance of ALT with a 2-layer g is close to RandConv, and is greater than all previous baselines for higher depth of the network. We do not see a clear correlation across datasets between the number of layers and the domain generalization performance. Investigating the dynamics of model capac-

Variable	Digits	PACS	Office-Home
f architecture	DigitNet [4]	ResNet18 [2]	ResNet50 [2]
g architecture		$\{\text{conv}_{kernel=3,stride=1,padding=1,leakyReLU_{p=0.2}}\} \times 4$	
ρ_0, ϕ_0		Kaiming Normal Initialization [1]	
T	10000	2000	2000
T_{pre}	1250	400	400
η	1e-4	0.004	0.004
m_{adv}	10	10	10
η_{adv}	5e-6	5e-5	5e-5
w_r	1.0	1.0	1.0
λ_{KL}	0.75	0.75	0.75

Table 1. Training settings and hyper-parameters for experiments on each benchmark.

	FCN Number of Layers				
Benchmark	2	3	4	5	6
Digits	72.75	73.74	74.10	73.87	74.15
PACS	63.40	63.92	64.41	64.20	63.78
OfficeHome	59.67	59.56	59.79	59.42	59.45

Table 2. Effect of the depth (number of convolutional layers) of the adversity network g on average domain generalization on all three benchmarks.

ity of the adversary network and how it may affect domain generalization, is an interesting direction for future work.

It may be possible that more complex generative architectures (i.e. greater complexity of transformations) may be needed for larger domain shift is larger, to model diversity and adversity for a given source domain. Thus the choice of architecture for g is an interesting direction; nevertheless, in this paper we show that the simple fully convolutional architecture gives us performance boosts in all three datasets.

We believe that ideas presented in this paper, although evaluated on image classification, have the potential of being widely applicable to many other vision tasks for domain generalization. They may also be applied to other application areas such as audio or text, where the transformation function g may take different forms.

References

[1] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *Proceedings of the IEEE international conference on computer vision*, pages 1026–1034, 2015. 2

[2] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, June 27-30, 2016*, pages 770–778. IEEE Computer Society, 2016. 2

[3] Dan Hendrycks, Norman Mu, Ekin Dogus Cubuk, Barret Zoph, Justin Gilmer, and Balaji Lakshminarayanan. Augmix: A simple data processing method to improve robustness and uncertainty. In *8th International Conference on Learning Representations, ICLR 2020, Addis Ababa, Ethiopia, April 26-30, 2020*. OpenReview.net, 2020. 1

[4] Riccardo Volpi, Hongseok Namkoong, Ozan Sener, John C. Duchi, Vittorio Murino, and Silvio Savarese. Generalizing to unseen domains via adversarial data augmentation. In *Advances in Neural Information Processing Systems*, pages 5339–5349, 2018. 2

[5] Zhenlin Xu, Deyi Liu, Junlin Yang, Colin Raffel, and Marc Niethammer. Robust and generalizable visual representation learning via random convolutions. In *International Conference on Learning Representations*, 2020. 1

Method	Photo*	Art-Painting	Cartoon	Sketch	Target Avg.	PACS Avg.
RandConv	96.407 ± 0.757	61.309 ± 2.316	37.577 ± 2.257	50.463 ± 9.018	49.783 ± 4.255	61.439 ± 3.217
AugMix	**99.532** ± 0.438	68.633 ± 0.950	33.788 ± 1.205	36.304 ± 2.801	46.242 ± 1.122	59.564 ± 0.930
RandConv + AugMix	98.363 ± 0.438	65.527 ± 3.060	39.300 ± 6.237	40.901 ± 5.073	48.576 ± 4.031	61.023 ± 3.001
ALT_{g-only}	99.064 ± 0.286	**68.770** ± 0.932	43.387 ± 1.142	50.832 ± 2.937	54.330 ± 1.078	65.513 ± 0.757
$\text{ALT}_{RandConv}$	98.947 ± 0.234	68.740 ± 0.702	40.828 ± 2.537	**56.024** ± 2.009	**55.197** ± 0.498	66.135 ± 0.330
ALT_{AugMix}	99.298 ± 0.438	68.506 ± 0.836	**43.507** ± 2.615	53.271 ± 4.149	55.094 ± 1.876	**66.145** ± 1.387

Table 3. SSDG performance on PACS for the P→ACS setting. *Source Domain. **bold**: best result.

Method	Photo	Art Painting*	Cartoon	Sketch	Target Avg.	PACS Avg.
RandConv	87.281 ± 0.796	85.437 ± 0.532	61.143 ± 2.752	60.519 ± 4.050	69.648 ± 2.152	73.595 ± 1.582
AugMix	**95.317** ± 0.422	**93.077** ± 1.276	64.061 ± 0.361	55.027 ± 2.195	71.469 ± 0.637	76.871 ± 0.581
RandConv + AugMix	90.743 ± 0.781	90.481 ± 0.638	64.206 ± 2.238	62.515 ± 2.854	72.488 ± 1.731	76.986 ± 1.177
ALT_{g-only}	94.934 ± 0.269	91.058 ± 0.720	63.524 ± 1.821	63.813 ± 2.249	74.090 ± 1.086	78.332 ± 0.845
$\text{ALT}_{RandConv}$	93.593 ± 0.328	92.596 ± 1.036	64.044 ± 0.635	65.991 ± 1.130	74.543 ± 0.537	79.056 ± 0.609
ALT_{AugMix}	93.174 ± 0.437	91.442 ± 0.638	**65.683** ± 1.656	**68.226** ± 2.453	**75.694** ± 1.214	**79.631** ± 0.856

Table 4. SSDG performance on PACS for the A→PCS setting. *Source Domain. **bold**: best result.

Method	Photo	Art-Painting	Cartoon*	Sketch	Target Avg.	PACS Avg.
RandConv	73.677 ± 1.814	57.051 ± 1.764	91.66 ± 0.876	72.855 ± 2.314	67.861 ± 1.550	73.810 ± 1.317
AugMix	84.599 ± 0.997	68.281 ± 2.085	**96.287** ± 0.940	71.097 ± 0.609	74.659 ± 1.088	80.066 ± 0.897
RandConv + AugMix	78.790 ± 0.975	65.400 ± 1.611	93.840 ± 1.020	71.285 ± 2.730	71.825 ± 1.315	77.329 ± 1.105
ALT_{g-only}	84.575 ± 1.047	68.867 ± 2.126	94.768 ± 0.43	74.421 ± 0.441	75.954 ± 1.119	80.658 ± 0.929
$\text{ALT}_{RandConv}$	83.916 ± 0.51	68.086 ± 1.901	95.190 ± 0.686	**74.487** ± 0.505	75.496 ± 0.799	80.420 ± 0.644
ALT_{AugMix}	**85.964** ± 1.098	**71.943** ± 1.234	94.599 ± 0.560	74.172 ± 0.752	**77.360** ± 0.734	**81.670** ± 0.667

Table 5. SSDG performance on PACS for the C→PAS setting. *Source Domain. **bold**: best result.

Method	Photo	Art-Painting	Cartoon	Sketch*	Target Avg.	PACS Avg.
RandConv	46.132 ± 4.879	**52.168** ± 1.623	**63.942** ± 2.219	94.264 ± 0.673	**54.081** ± 1.959	**64.126** ± 1.465
AugMix	46.731 ± 2.916	37.852 ± 1.878	58.575 ± 1.747	94.221 ± 0.711	47.719 ± 1.723	59.345 ± 1.268
RandConv + AugMix	**54.359** ± 0.819	46.074 ± 2.709	61.246 ± 1.245	94.171 ± 0.582	53.893 ± 0.945	63.963 ± 0.787
ALT_{g-only}	49.305 ± 2.775	39.658 ± 3.423	61.109 ± 1.853	94.573 ± 0.466	50.024 ± 2.408	61.161 ± 1.726
$\text{ALT}_{RandConv}$	51.305 ± 0.866	41.787 ± 1.174	62.773 ± 1.089	**94.724** ± 0.527	51.955 ± 0.791	62.647 ± 0.571
ALT_{AugMix}	49.078 ± 2.072	40.186 ± 2.494	62.901 ± 0.358	94.271 ± 0.624	50.721 ± 1.414	61.609 ± 1.103

Table 6. SSDG performance on PACS for the S→PAC setting. *Source Domain. **bold**: best result.

Method	**Real**⋆	**Art**	**Clipart**	**Product**	**Target Avg.**	**Office-Home Avg.**
RandConv	83.028 $\pm$0.067	59.021 $\pm$0.916	47.269 $\pm$1.251	72.172 $\pm$0.418	59.487 $\pm$0.792	65.372 $\pm$1.096
AugMix	87.294 $\pm$1.21	64.101 $\pm$0.882	47.564 $\pm$0.158	75.956 $\pm$0.32	62.54 $\pm$0.345	68.729 $\pm$0.490
RandConv + AugMix	81.514 $\pm$0.515	59.167 $\pm$0.722	48.180 $\pm$1.024	71.166 $\pm$0.445	59.504 $\pm$0.256	65.007 $\pm$0.226
ALT$_{g-only}$	86.514 $\pm$0.622	64.622 $\pm$0.490	**53.327** $\pm$0.344	76.276 $\pm$0.117	64.742 $\pm$0.122	70.185 $\pm$0.138
ALT$_{RandConv}$	**87.477** $\pm$1.042	**64.879** $\pm$0.439	53.097 $\pm$0.554	76.066 $\pm$0.447	64.681 $\pm$0.290	**70.380** $\pm$0.312
ALT$_{AugMix}$	86.560 $\pm$0.980	64.860 $\pm$0.267	53.271 $\pm$0.799	**76.286** $\pm$0.347	**64.806** $\pm$0.327	70.244 $\pm$0.461

Table 7. SSDG performance on Office-Home for the R→ACP setting. ⋆Source Domain. **bold**: best result.

Method	**Real**	**Art**⋆	**Clipart**	**Product**	**Target Avg.**	**Office-Home Avg.**
RandConv	66.915 $\pm$1.069	72.428 $\pm$2.066	42.387 $\pm$1.405	55.045 $\pm$1.547	54.782 $\pm$1.297	59.194 $\pm$1.427
AugMix	**71.887** $\pm$0.432	**80.494** $\pm$1.342	45.314 $\pm$0.768	**61.882** $\pm$0.382	59.694 $\pm$0.427	**64.894** $\pm$0.369
RandConv + AugMix	65.620 $\pm$0.632	71.852 $\pm$1.758	42.606 $\pm$1.026	54.434 $\pm$0.774	54.220 $\pm$0.617	58.628 $\pm$0.862
ALT$_{g-only}$	71.193 $\pm$0.308	78.930 $\pm$1.146	47.340 $\pm$0.331	61.151 $\pm$0.561	59.895 $\pm$0.283	64.654 $\pm$0.259
ALT$_{RandConv}$	71.754 $\pm$0.286	78.025 $\pm$1.181	**48.328** $\pm$0.787	61.186 $\pm$0.429	**60.423** $\pm$0.280	64.823 $\pm$0.474
ALT$_{AugMix}$	71.122 $\pm$0.540	79.095 $\pm$1.634	48.058 $\pm$0.632	61.156 $\pm$0.813	60.112 $\pm$0.590	64.858 $\pm$0.518

Table 8. SSDG performance on Office-Home for the A→RCP setting. ⋆Source Domain. **bold**: best result.

Method	**Real**	**Art**	**Clipart**⋆	**Product**	**Target Avg.**	**Office-Home Avg.**
RandConv	58.944 $\pm$0.521	44.741 $\pm$0.714	80.320 $\pm$1.073	56.211 $\pm$1.141	53.299 $\pm$0.711	60.054 $\pm$0.771
AugMix	62.244 $\pm$0.526	49.309 $\pm$0.879	**81.510** $\pm$0.885	**58.939** $\pm$0.584	56.831 $\pm$0.530	**63.000** $\pm$0.454
RandConv + AugMix	57.914 $\pm$0.730	43.698 $\pm$0.511	77.986 $\pm$1.087	55.040 $\pm$0.683	52.217 $\pm$0.249	58.660 $\pm$0.417
ALT$_{g-only}$	61.968 $\pm$0.849	49.977 $\pm$0.987	80.320 $\pm$1.073	58.779 $\pm$0.743	56.908 $\pm$0.808	62.761 $\pm$0.733
ALT$_{RandConv}$	**62.264** $\pm$0.560	50.133 $\pm$0.956	80.732 $\pm$0.637	58.819 $\pm$0.558	**57.072** $\pm$0.539	62.987 $\pm$0.455
ALT$_{AugMix}$	61.841 $\pm$0.382	**50.426** $\pm$1.070	80.824 $\pm$0.510	58.839 $\pm$0.559	57.035 $\pm$0.580	62.982 $\pm$0.526

Table 9. SSDG performance on Office-Home for the C→RAP setting. ⋆Source Domain. **bold**: best result.

Method	**Real**	**Art**	**Clipart**	**Product**⋆	**Target Avg.**	**Office-Home Avg.**
RandConv	66.318 $\pm$0.240	43.524 $\pm$0.664	43.365 $\pm$1.058	90.135 $\pm$0.643	51.069 $\pm$0.607	60.836 $\pm$0.372
AugMix	**71.515** $\pm$0.706	**50.041** $\pm$0.688	42.596 $\pm$0.619	**91.622** $\pm$0.263	54.717 $\pm$0.518	63.943 $\pm$0.453
RandConv + AugMix	65.523 $\pm$0.753	43.240 $\pm$1.454	41.710 $\pm$0.621	89.459 $\pm$0.785	50.158 $\pm$0.900	59.983 $\pm$0.865
ALT$_{g-only}$	70.082 $\pm$0.532	48.842 $\pm$0.648	46.877 $\pm$0.552	91.306 $\pm$0.544	55.267 $\pm$0.302	64.277 $\pm$0.171
ALT$_{RandConv}$	70.530 $\pm$0.359	49.208 $\pm$0.418	47.025 $\pm$0.498	91.577 $\pm$0.506	55.588 $\pm$0.300	64.585 $\pm$0.212
ALT$_{AugMix}$	70.637 $\pm$0.301	49.318 $\pm$1.008	**47.554** $\pm$0.458	91.396 $\pm$0.798	**55.837** $\pm$0.383	**64.726** $\pm$0.361

Table 10. SSDG performance on Office-Home for the P→RAC setting. ⋆Source Domain. **bold**: best result.

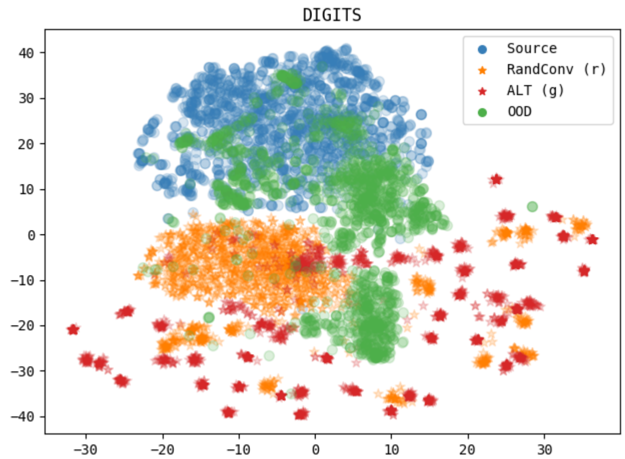

Figure 1. tSNE plot showing the discrepancy between the source distribution and the out-of-distribution datasets for the Digits benchmark.

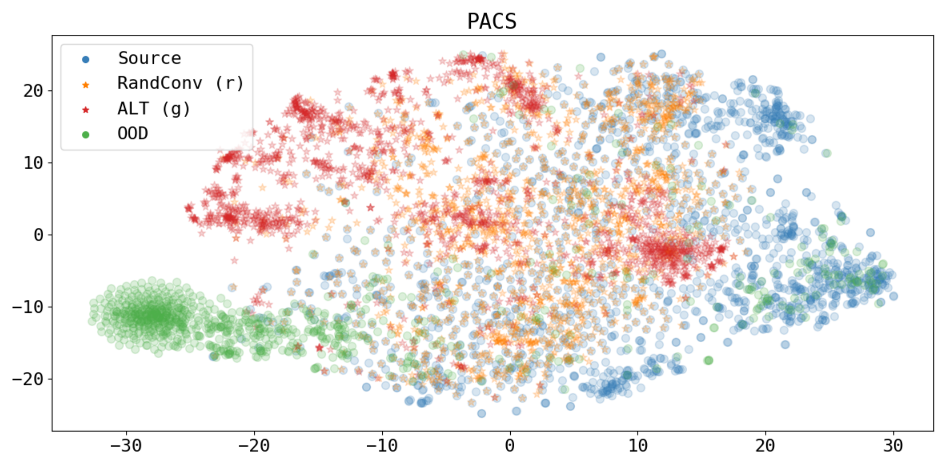

Figure 2. tSNE plot showing the discrepancy between the source distribution and the out-of-distribution datasets for the PACS benchmark.

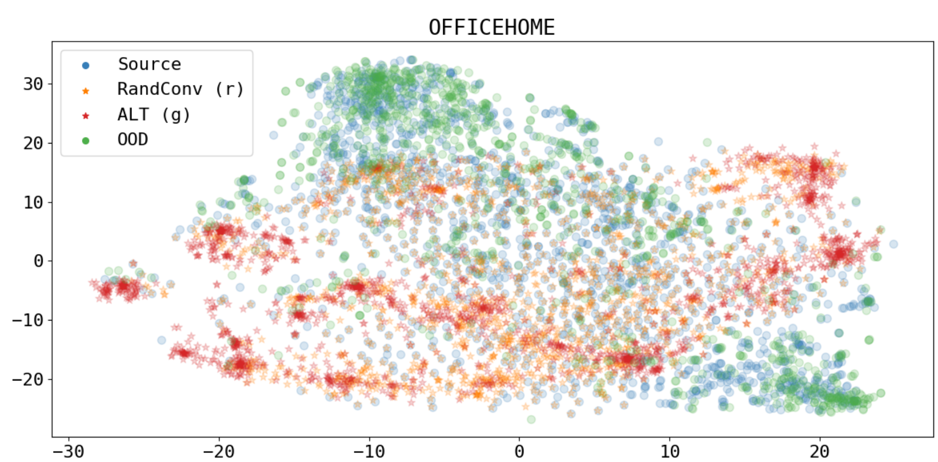

Figure 3. tSNE plot showing the discrepancy between the source distribution and the out-of-distribution datasets for the OfficeHome benchmark.

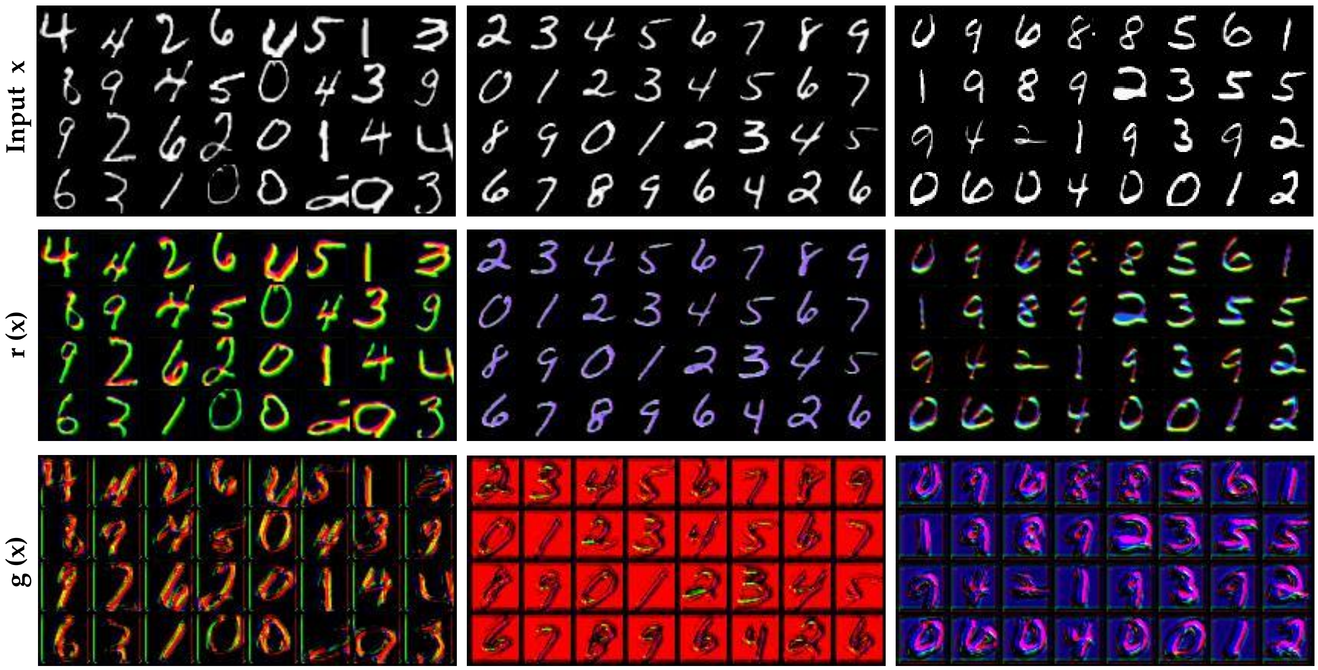

Figure 4. Digits: Comparison of images transformed by RandConv and ALT$_{RandConv}$ with MNIST10k as source dataset.

Figure 5. PACS: Comparison of images transformed by RandConv and ALT$_{RandConv}$ with *Photo* as source dataset.

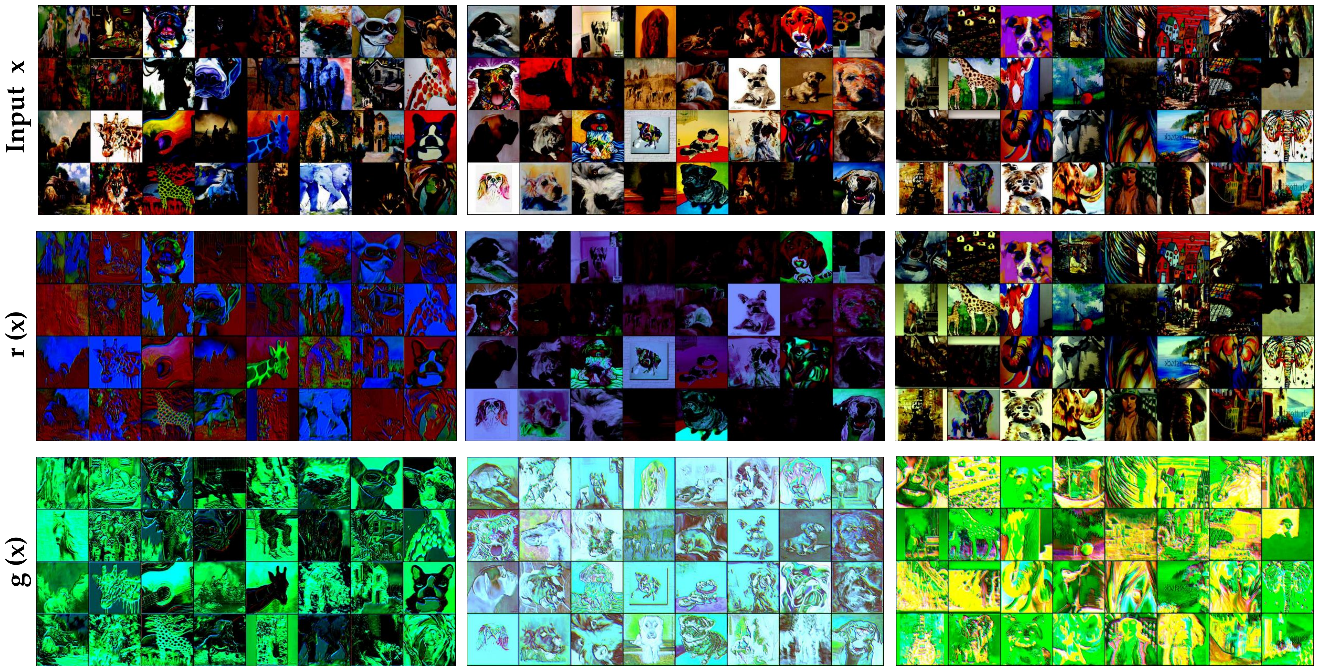

Figure 6. PACS: Comparison of images transformed by RandConv and ALT$_{RandConv}$ with *Art-Painting* as source dataset.

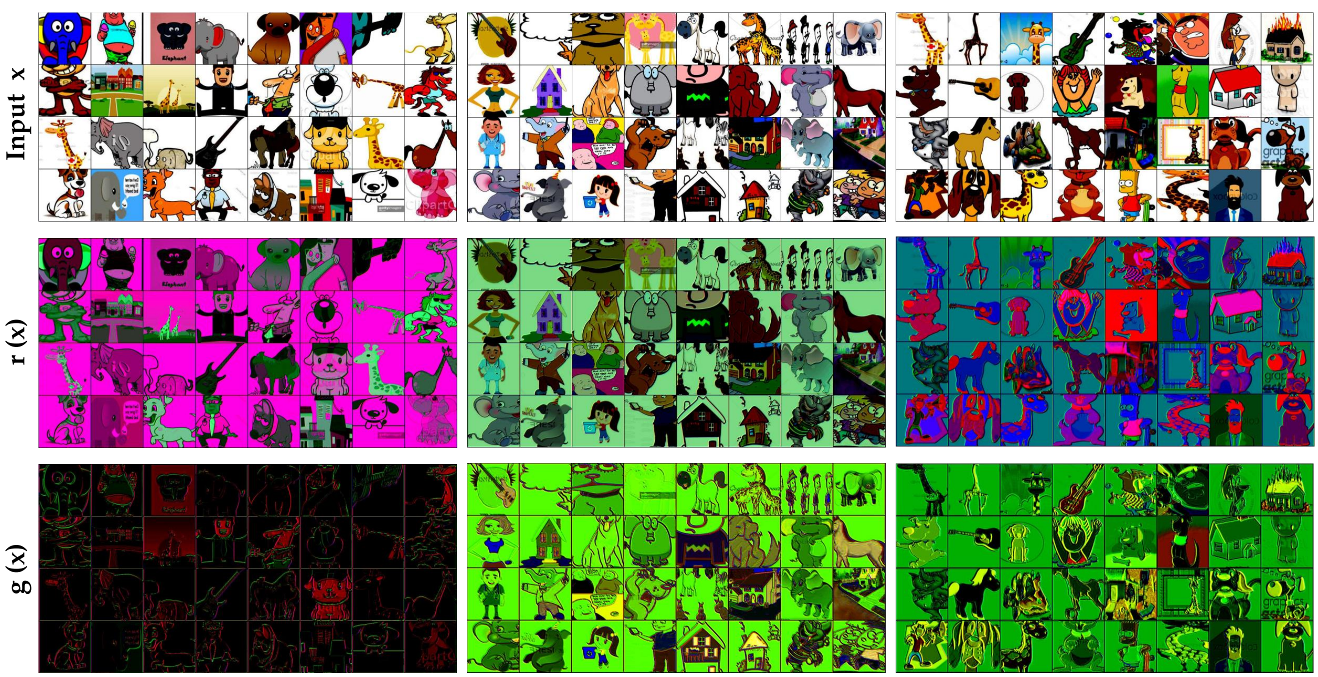

Figure 7. PACS: Comparison of images transformed by RandConv and ALT$_{RandConv}$ with *Cartoon* as source dataset.

Figure 8. PACS: Comparison of images transformed by RandConv and ALT$_{RandConv}$ with *Sketch* as source dataset.

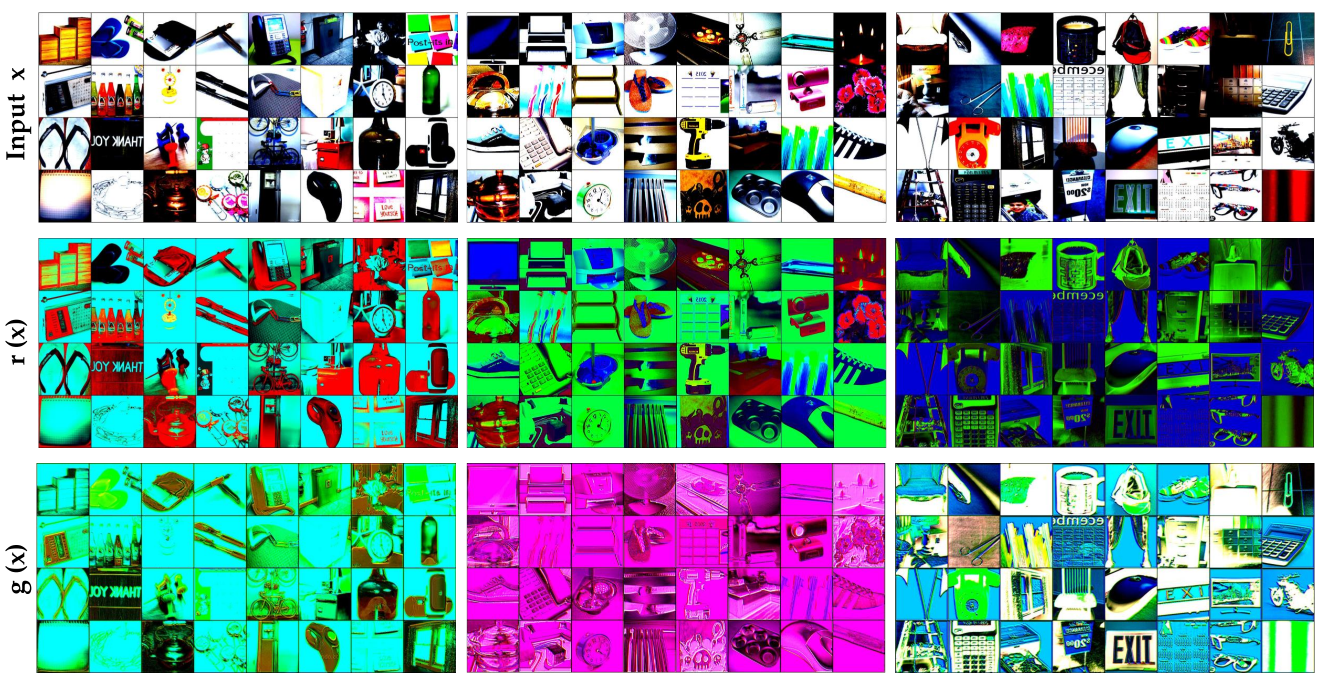

Figure 9. Office-Home: Comparison of images transformed by RandConv and ALT$_{RandConv}$ with *Real* as source dataset.

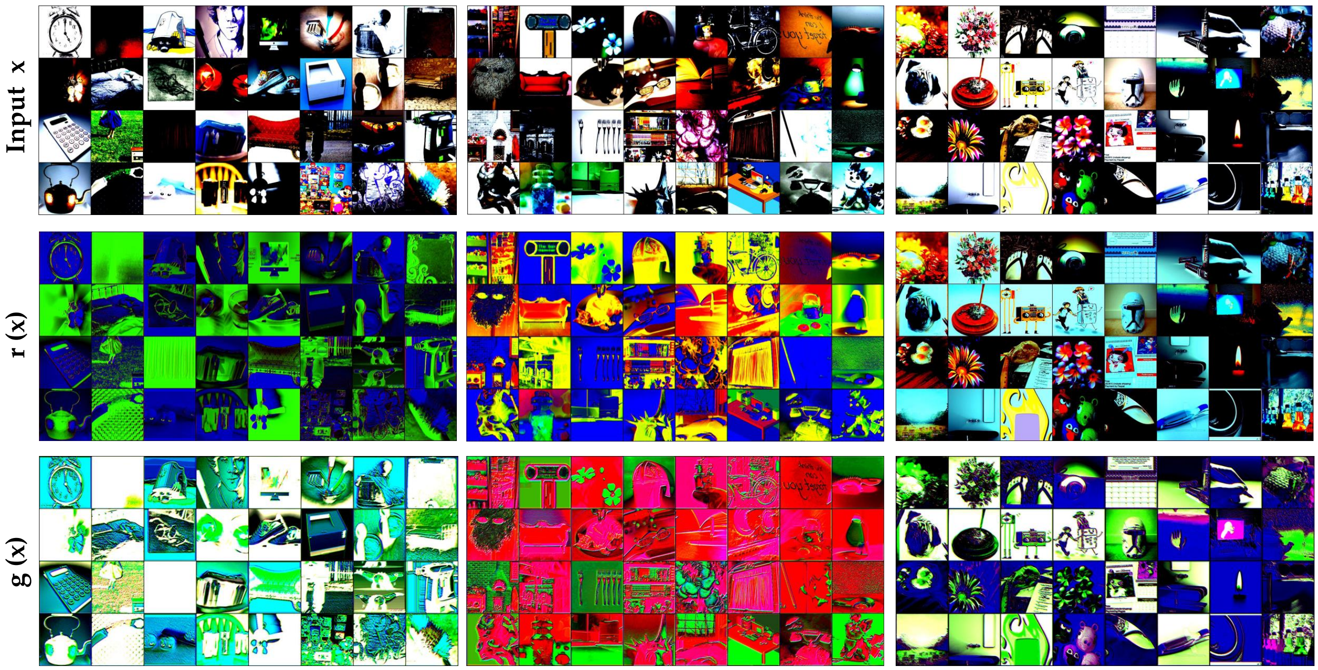

Figure 10. Office-Home: Comparison of images transformed by RandConv and ALT$_{RandConv}$ with *Art* as source dataset.

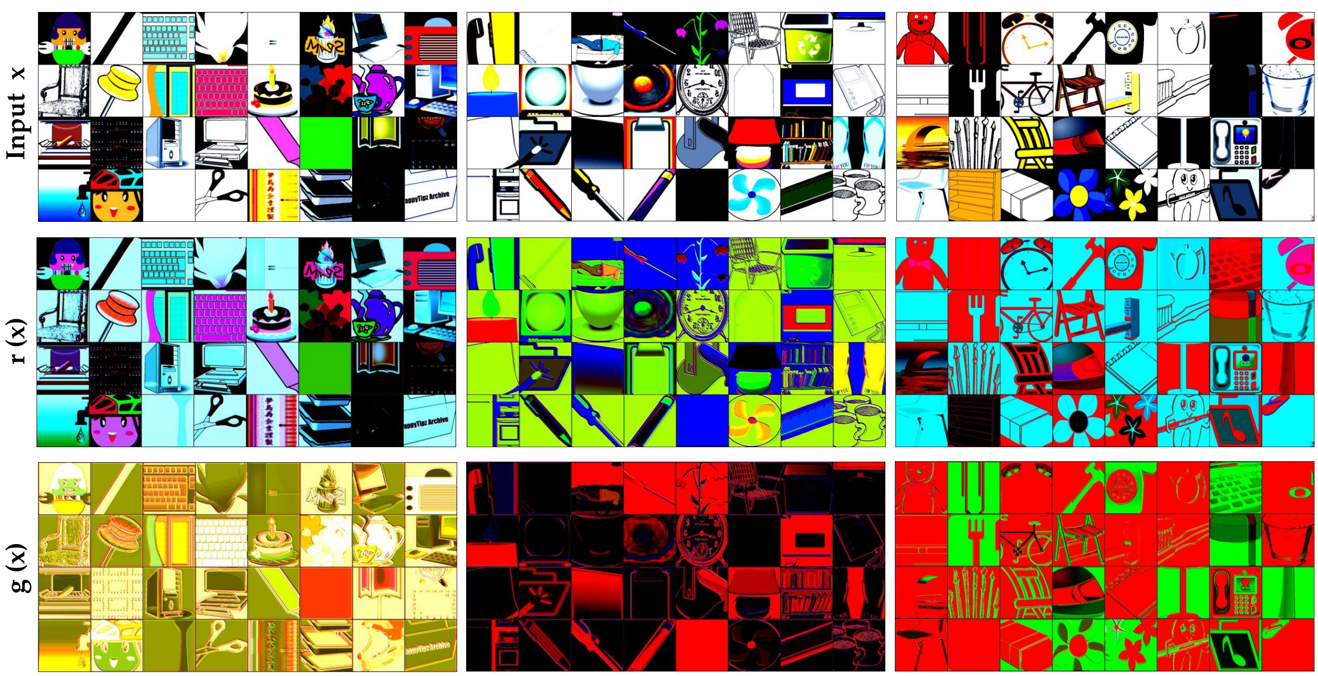

Figure 11. Office-Home: Comparison of images transformed by RandConv and ALT$_{RandConv}$ with *Clipart* as source dataset.

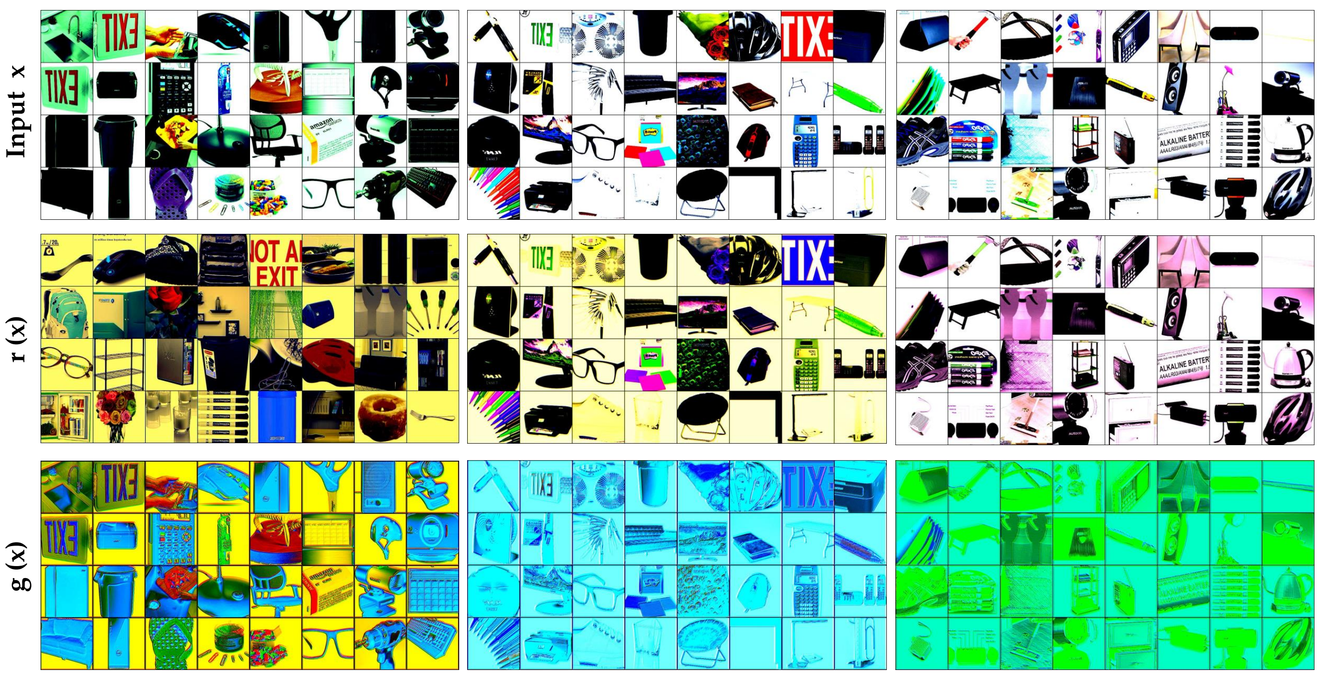

Figure 12. Office-Home: Comparison of images transformed by RandConv and ALT$_{RandConv}$ with *Product* as source dataset.