Appendix

ImageNet-BG: A Toolkit and Dataset for Evaluating Vision Model Robustness Against Background Variations

Łukasz Piekarek Kamil Szyc
Faculty of Information and Communication Technology
Wrocław University of Science and Technology, Wrocław, Poland
lukaszpiekarekt5346@gmail.com, kamil.szyc@pwr.edu.pl

ImageNet-GB and some other supplementary materials are available in our Google Drive. The code is publicly available at github.com/TrustAIRiders/ICCVW-2025-ImageNet-BG

A. ImageNet-BG

The dataset consists of 434,121 instances, including all qualities and the unaltered images. All included images use standard data formats and are normalized to a single size, so no additional preprocessing is needed to use the dataset. Labeling is achieved by ordering images in a directory hierarchy as follows: quality - background type - background image - class. (e.g., v_good/real/beach)

To make the data structures more humanly readable, the star rating system was replaced with readable words: N/A - bad, ⋆ - questionable, ⋆⋆ - ok, ⋆ ⋆ ⋆ - good, ⋆ ⋆ ⋆⋆ - v_good.

The dataset has no training/validation split as it is meant primarily for testing. The names of backgrounds contain legacy suffixes indicating the acquisition method of each background: ah - method 1, gah - method 2, and sau - method 3.

Additional materials, such as HAQA, conform to the same naming scheme. HAQA is a CSV spreadsheet with 8 columns: *Class Label, Class Name, Overall Quality, Detail, Crop, Size, Distinctiveness, and Notes*. The first 2 columns are for class identification, providing the ImageNet-compliant and human-readable labels. The next 5 columns relate to the quality assessment outlined in the main paper, with *Overall Quality* being the final score assigned to a class. The *Notes* column also provides additional annotator commentary for unusual cases. The backgrounds used are an additional resource and are split by background category.

B. Toolkit

The toolkit is a collection of Python scripts enabling the reproduction of our dataset generation process. Usage instructions are included with the code as a text file. It includes the routines necessary for background selection and substitution based on the ImageNet-S dataset.

A separate testing environment is available alongside the toolkit, enabling easy reproduction of the results showcased in the paper.

C. Extended Results

Tables 1 and 2 are equivalent to Table 1 in the main paper but configured for different HAQA setups. They provide a more comprehensive analysis of model performance across various configurations, ensuring the robustness of our conclusions.

The table 3 further distinguishes between different approaches for Real background acquisition. Despite these variations in background selection, the results indicate no significant differences between them.

Tables 1, 2, and 3 reinforce the primary conclusions drawn from Table 1 in the main paper, validating the consistency and reliability of our findings across different HAQA.

Tables 4, 5, 6 and 7 are equivalent to table 2 in the main paper but explore a variety of models (tables 4 and 7) within the HAQA (tables 5 and 6). Despite the different models and configurations, the primary conclusions align with those presented in the main paper, reaffirming the consistency of our findings.

Moreover, table 7 provides detailed results for specific models, illustrating that the challenges associated with certain classes are more influenced by the robustness based on the training dataset rather than the model architecture itself. This highlights the importance of dataset composition in achieving robust performance.

Table 1. This table is equivalent to Table 1 in the main paper, presenting model performance for all ImageNet-BG classes (based on HAQA).

Model	**Acc Unaltered**	**Acc Abstract**	**Acc Real**
AlexNet	58.1	37.9 Δ -20.2	32.7 Δ -25.4
VGG 19	76.3	65.1 Δ -11.2	60.2 Δ -16.1
Inception V3	79.0	67.0 Δ -12.0	63.7 Δ -15.3
ResNet 152	80.6	**69.5 Δ -11.1**	**66.4 Δ -14.2**
ShuffleNet V2	74.5	**62.9 Δ -11.7**	**57.7 Δ -16.8**
MobileNet V3	75.8	60.9 Δ -14.9	51.5 Δ -24.3
RegNetY 800MF	78.5	65.0 Δ -13.5	59.2 Δ -19.3
EfficientNet B3	82.8	73.4 Δ -9.5	68.4 Δ -14.4
ConvNeXt L	84.8	**79.5 Δ -5.3**	**75.3 Δ -9.5**
EfficientNet V2M	84.3	76.2 Δ -8.2	72.3 Δ -12.0
ViT L	81.8	70.4 Δ -11.4	63.8 Δ -18.0
MaxViT T	84.5	**76.4 Δ -8.1**	**72.8 Δ -11.7**
SwinV2 B	83.6	75.1 Δ -8.6	70.8 Δ -12.9

Table 2. This table is equivalent to Table 1 in the main paper, presenting model performance only for ImageNet-BG classes with ⋆ ⋆ ⋆⋆ in HAQA).

Model	**Acc Unaltered**	**Acc Abstract**	**Acc Real**
AlexNet	61.5	47.1 Δ -14.3	39.1 Δ -22.4
VGG 19	79.5	73.5 Δ -5.9	68.4 Δ -11.0
Inception V3	82.5	75.6 Δ -6.9	71.6 Δ -10.9
ResNet 152	84.4	**78.3 Δ -6.1**	**75.2 Δ -9.3**
ShuffleNet V2	77.9	**72.0 Δ -5.9**	**66.4 Δ -11.4**
MobileNet V3	79.5	71.3 Δ -8.2	60.5 Δ -19.0
RegNetY 800MF	82.1	74.7 Δ -7.4	67.8 Δ -14.3
EfficientNet B3	86.3	81.7 Δ -4.6	77.1 Δ -9.2
ConvNeXt L	87.2	**85.7 Δ -1.6**	**82.3 Δ -4.9**
EfficientNet V2M	87.0	82.5 Δ -4.5	79.2 Δ -7.8
ViT L	85.4	79.1 Δ -6.2	73.0 Δ -12.3
MaxViT T	87.5	**84.1 Δ -3.4**	**80.9 Δ -6.6**
SwinV2 B	86.3	82.1 Δ -4.2	78.6 Δ -7.7

Table 3. This table is equivalent to Table 1 in the main paper, presenting model performance only for ImageNet-BG classes with ⋆ ⋆ ⋆ or ⋆ ⋆ ⋆⋆ in HAQA, distinguishing between different approaches for selecting Real backgrounds.

Model	**Acc Unaltered**	**Acc Real (Approach 1)**	**Acc Real (Approach 2)**	**Acc Real (Approach 3)**
AlexNet	59.7	36.1 Δ -23.6	36.7 Δ -23.0	34.5 Δ -25.1
VGG 19	77.8	65.0 Δ -12.8	65.7 Δ -12.1	63.5 Δ -14.3
Inception V3	80.5	68.2 Δ -12.3	68.8 Δ -11.7	67.7 Δ -12.8
ResNet 152	82.1	**71.0 Δ -11.0**	**71.4 Δ -10.7**	**70.6 Δ -11.5**
ShuffleNet V2	75.9	**62.8 Δ -13.1**	63.0 Δ -12.9	**61.3 Δ -14.6**
MobileNet V3	77.2	56.2 Δ -21.0	56.4 Δ -20.8	54.5 Δ -22.7
RegNetY 800MF	80.0	63.8 Δ -16.2	**64.0 Δ -16.1**	62.7 Δ -17.3
EfficientNet B3	84.0	73.4 Δ -10.6	73.6 Δ -10.4	72.0 Δ -12.0
ConvNeXt L	85.9	**79.7 Δ -6.1**	**79.7 Δ -6.1**	**78.9 Δ -7.0**
EfficientNet V2M	85.5	76.3 Δ -9.2	77.4 Δ -8.0	76.0 Δ -9.5
ViT L	83.2	69.6 Δ -13.6	68.3 Δ -14.9	67.8 Δ -15.4
MaxViT T	85.7	**77.2 Δ -8.6**	**77.9 Δ -7.9**	**77.0 Δ -8.7**
SwinV2 B	84.8	74.7 Δ -10.1	75.7 Δ -9.1	74.5 Δ -10.4

Table 4. This table is equivalent to Table 2 in the main paper but includes performance metrics for all models and classes based on HAQA.

Class	**Unaltered**	**Abstract**	**Real**
n03733131 ⋆ maypole	98.4	5.4 Δ -93.0	10.2 Δ -88.2
n02817516 ⋆ bearskin	95.3	11.1 Δ -84.2	17.0 Δ -78.3
n04118538 ⋆⋆ rugby ball	95.6	18.0 Δ -77.6	15.4 Δ -80.2
n03929855 ⋆ pickelhaube	84.6	8.2 Δ -76.4	7.2 Δ -77.4
n04162706 ⋆ seat belt	85.1	9.9 Δ -75.2	10.5 Δ -74.6
n04540053 ⋆⋆ volleyball	93.8	18.8 Δ -75.0	19.2 Δ -74.6
n02802426 ⋆⋆ basketball	96.4	23.3 Δ -73.1	21.9 Δ -74.5
n03868242 ⋆ oxcart	83.4	8.2 Δ -75.2	11.5 Δ -71.9
n03814639 ⋆ neck brace	73.7	2.1 Δ -71.6	0.4 Δ -73.3
n03538406 ⋆⋆ horse cart	88.7	14.4 Δ -74.4	19.8 Δ -68.9
n04465501 ⋆ ⋆ ⋆ tractor	71.4	87.9 Δ +16.5	83.0 Δ +11.6
n04591713 ⋆ ⋆ ⋆ wine bottle	36.2	56.2 Δ +20.0	44.5 Δ +8.3
n04399382 ⋆⋆ teddy	41.6	61.5 Δ +19.8	50.7 Δ +9.1
n03633091 ⋆ ⋆ ⋆ ladle	28.2	51.1 Δ +22.9	35.8 Δ +7.6
n07768694 ⋆ ⋆ ⋆⋆ pomegranate	65.4	83.8 Δ +18.4	77.8 Δ +12.4
n03594734 ⋆ ⋆ ⋆ jean	30.4	53.4 Δ +23.0	41.1 Δ +10.7
n03832673 ⋆⋆ notebook	16.9	41.8 Δ +24.8	26.4 Δ +9.4
n03691459 ⋆ ⋆ ⋆ loudspeaker	62.2	81.1 Δ +18.9	80.1 Δ +17.8
n02971356 ⋆ ⋆ ⋆ carton	41.0	67.2 Δ +26.2	51.6 Δ +10.6
n03532672 ⋆ ⋆ ⋆ hook	35.9	67.3 Δ +31.4	44.7 Δ +8.8

Table 5. This table is equivalent to Table 2 in the main paper but includes performance metrics for selected models and only for classes with ⋆ ⋆ ⋆⋆ in HAQA.

Class	Unaltered	Abstract	Real
n02910353 ⋆ ⋆ ⋆⋆ buckle	83.3	56.5 Δ -26.9	31.5 Δ -51.9
n04606251 ⋆ ⋆ ⋆⋆ wreck	100.0	65.4 Δ -34.6	57.4 Δ -42.6
n04579432 ⋆ ⋆ ⋆⋆ whistle	100.0	74.4 Δ -25.6	57.2 Δ -42.8
n07584110 ⋆ ⋆ ⋆⋆ consomme	90.3	62.2 Δ -28.1	51.7 Δ -38.6
n02277742 ⋆ ⋆ ⋆⋆ ringlet	94.4	65.6 Δ -28.9	65.1 Δ -29.3
n02730930 ⋆ ⋆ ⋆⋆ apron	90.3	60.1 Δ -30.2	63.8 Δ -26.5
n04026417 ⋆ ⋆ ⋆⋆ purse	67.8	38.9 Δ -28.9	40.4 Δ -27.4
n03062245 ⋆ ⋆ ⋆⋆ cocktail shaker	96.3	69.4 Δ -26.9	68.5 Δ -27.8
n02981792 ⋆ ⋆ ⋆⋆ catamaran	87.5	63.5 Δ -24.0	57.2 Δ -30.3
n02169497 ⋆ ⋆ ⋆⋆ leaf beetle	85.2	64.4 Δ -20.8	52.2 Δ -33.0
n03709823 ⋆ ⋆ ⋆⋆ mailbag	52.0	62.3 Δ +10.3	59.8 Δ +7.8
n03938244 ⋆ ⋆ ⋆⋆ pillow	73.5	86.5 Δ +12.9	78.9 Δ +5.3
n02109961 ⋆ ⋆ ⋆⋆ Eskimo dog	41.7	51.7 Δ +10.0	51.9 Δ +10.2
n04398044 ⋆ ⋆ ⋆⋆ teapot	65.7	76.2 Δ +10.4	75.7 Δ +10.0
n03063689 ⋆ ⋆ ⋆⋆ coffeepot	62.5	69.4 Δ +6.9	78.8 Δ +16.3
n06596364 ⋆ ⋆ ⋆⋆ comic book	63.1	77.1 Δ +14.0	74.3 Δ +11.2
n03063599 ⋆ ⋆ ⋆⋆ coffee mug	66.7	79.7 Δ +13.0	80.1 Δ +13.5
n01773797 ⋆ ⋆ ⋆⋆ garden spider	26.7	40.0 Δ +13.3	40.8 Δ +14.1
n07768694 ⋆ ⋆ ⋆⋆ pomegranate	70.8	87.2 Δ +16.4	83.4 Δ +12.5
n04505470 ⋆ ⋆ ⋆⋆ typewriter keyboard	54.2	80.2 Δ +26.0	74.3 Δ +20.1

Table 6. This table is equivalent to Table 2 in the main paper but includes performance metrics for selected models and only for classes with ⋆⋆ in HAQA.

Class	Unaltered	Abstract	Real
n04118538 ⋆⋆ rugby ball	100.0	22.3 Δ -77.7	21.1 Δ -78.9
n03538406 ⋆⋆ horse cart	96.7	21.9 Δ -74.7	28.1 Δ -68.6
n03188531 ⋆⋆ diaper	87.2	26.6 Δ -60.6	10.5 Δ -76.7
n04209133 ⋆⋆ shower cap	84.5	20.2 Δ -64.3	12.8 Δ -71.7
n02802426 ⋆⋆ basketball	98.9	32.8 Δ -66.1	29.7 Δ -69.2
n03770439 ⋆⋆ miniskirt	73.5	10.5 Δ -63.0	2.2 Δ -71.3
n04019541 ⋆⋆ puck	97.4	33.7 Δ -63.8	29.1 Δ -68.3
n02951358 ⋆⋆ canoe	90.5	22.6 Δ -67.9	26.6 Δ -63.9
n04540053 ⋆⋆ volleyball	95.6	33.3 Δ -62.2	27.4 Δ -68.1
n03494278 ⋆⋆ harmonica	98.5	36.0 Δ -62.5	34.3 Δ -64.1
n01740131 ⋆⋆ night snake	41.7	50.0 Δ +8.3	29.0 Δ -12.7
n03249569 ⋆⋆ drum	78.6	80.4 Δ +1.8	75.4 Δ -3.2
n04090263 ⋆⋆ rifle	48.6	51.7 Δ +3.1	44.1 Δ -4.5
n03958227 ⋆⋆ plastic bag	62.2	63.6 Δ +1.4	60.4 Δ -1.9
n04507155 ⋆⋆ umbrella	60.8	64.6 Δ +3.8	57.4 Δ -3.4
n04154565 ⋆⋆ screwdriver	30.8	44.2 Δ +13.5	21.8 Δ -8.9
n03045698 ⋆⋆ cloak	47.6	55.4 Δ +7.7	52.4 Δ +4.8
n03000134 ⋆⋆ chainlink fence	53.2	67.5 Δ +14.3	69.9 Δ +16.7
n04399382 ⋆⋆ teddy	43.2	69.3 Δ +26.1	55.9 Δ +12.8
n03832673 ⋆⋆ notebook	20.7	52.0 Δ +31.3	35.0 Δ +14.4

Table 7. This table is equivalent to Table 2 in the main paper but includes performance metrics for all classes in HAQA, presented not as means but for selected models. For clarity, the only delta is shown as ΔAbstract/ΔReal.

ResNet 152	**ConvNeXt L**	**MaxVit T**
n13133613 ⋆ ear	n03733131 ⋆ maypole	n03733131 ⋆ maypole
Δ -100.0 / Δ -85.0	Δ -91.1 / Δ -86.7	Δ -92.9 / Δ -86.7
n03733131 ⋆ maypole	n03814639 ⋆ neck brace	n02817516 ⋆ bearskin
Δ -96.4 / Δ -86.7	Δ -77.1 / Δ -80.8	Δ -94.2 / Δ -85.1
n02817516 ⋆ bearskin	n03929855 ⋆ pickelhaube	n04162706 ⋆ seat belt
Δ -94.2 / Δ -85.9	Δ -75.0 / Δ -78.6	Δ -87.5 / Δ -85.2
n02802426 ⋆⋆basketball	n04162706 ⋆ seat belt	n03538406 ⋆⋆ horse cart
Δ -80.0 / Δ -77.8	Δ -81.2 / Δ -71.5	Δ -91.7 / Δ -75.3
n04209133 ⋆⋆ shower cap	n04118538 ⋆⋆ rugby ball	n03814639 ⋆ neck brace
Δ -76.8 / Δ -80.0	Δ -71.4 / Δ -76.0	Δ -81.2 / Δ -83.3
n04162706 ⋆ seat belt	n04209133 ⋆⋆ shower cap	n03868242 ⋆ oxcart
Δ -76.6 / Δ -80.0	Δ -66.1 / Δ -76.2	Δ -78.9 / Δ -78.2
n04118538 ⋆⋆ rugby ball	n03188531 ⋆⋆ diaper	n03929855 ⋆ pickelhaube
Δ -75.0 / Δ -81.2	Δ -55.8 / Δ -79.7	Δ -78.6 / Δ -78.6
n03929855 ⋆ pickelhaube	n03538406 ⋆⋆ horse cart	n04540053 ⋆⋆ volleyball
Δ -76.8 / Δ -78.6	Δ -66.7 / Δ -68.0	Δ -80.0 / Δ -73.6
n03868242 ⋆ oxcart	n04019541 ⋆⋆ puck	n04118538 ⋆⋆ rugby ball
Δ -77.6 / Δ -75.6	Δ -61.5 / Δ -72.6	Δ -75.0 / Δ -76.4
n03538406 ⋆⋆ horse cart	n03868242 ⋆ oxcart	n03134739 ⋆⋆ croquet ball
Δ -80.0 / Δ -71.8	Δ -67.1 / Δ -65.6	Δ -71.4 / Δ -75.5
n11939491 ⋆ ⋆ ⋆ daisy	n04409515 ⋆ ⋆ ⋆ tennis ball	n11939491 ⋆ ⋆ ⋆ daisy
Δ +23.6 / Δ +13.5	Δ +38.1 / Δ +13.0	Δ +27.8 / Δ +18.3
n03658185 ⋆ ⋆ ⋆⋆ letter opener	n03691459 ⋆ ⋆ ⋆loudspeaker	n04350905 ⋆ ⋆ ⋆suit
Δ +35.7 / Δ +1.4	Δ +25.0 / Δ +26.4	Δ +23.3 / Δ +24.2
n03063689 ⋆ ⋆ ⋆⋆ coffeepot	n04591713 ⋆ ⋆ ⋆ wine bottle	n04399382 ⋆⋆ teddy
Δ +12.5 / Δ +26.7	Δ +36.8 / Δ +17.8	Δ +29.5 / Δ +18.6
n02105412 ⋆ ⋆ ⋆ kelpie	n03063689 ⋆ ⋆ ⋆⋆ coffeepot	n03028079⋆ ⋆ ⋆ church
Δ +15.6 / Δ +24.2	Δ +25.0 / Δ +29.7	Δ +21.9 / Δ +28.5
n07768694 ⋆ ⋆ ⋆⋆p omegranate	n03832673 ⋆⋆ notebook	n04548362 ⋆ ⋆ ⋆⋆ wallet
Δ +23.4 / Δ +17.3	Δ +37.0 / Δ +18.5	Δ +30.0 / Δ +21.8
n04409515 ⋆ ⋆ ⋆ tennis ball	n04465501 ⋆ ⋆ ⋆ tractor	n03594734 ⋆ ⋆ ⋆ jean
Δ +29.8 / Δ +13.0	Δ +32.1 / Δ +27.8	Δ +31.2 / Δ +24.2
n04485082 ⋆ ⋆ ⋆ tripod	n03594734 ⋆ ⋆ ⋆ jean	n03832673 ⋆⋆ notebook
Δ +23.6 / Δ +19.6	Δ +37.5 / Δ +25.3	Δ +38.0 / Δ +20.7
n02971356 ⋆ ⋆ ⋆ carton	n01773797 ⋆ ⋆ ⋆⋆ garden spider	n03532672⋆ ⋆ ⋆ hook
Δ +35.0 / Δ +8.9	Δ +32.5 / Δ +31.3	Δ +36.1 / Δ +23.0
n03532672 ⋆ ⋆ ⋆ hook	n04525038 ⋆ velvet	n02971356 ⋆ ⋆ ⋆ carton
Δ +38.9 / Δ +13.7	Δ +50.0 / Δ +30.0	Δ +40.0 / Δ +37.3