EVET: Enhancing Visual Explanations of Deep Neural Networks Using Image Transformations
-Supplementary Materials-

Youngrock Oh
AI Advanced Research Lab, R&D Center, Samsung SDS
Seoul, Republic of Korea
y52.oh@samsung.com

Hyungsik Jung
jhyungsik89@gmail.com

Jeonghyung Park
jeonghyung.park@gmail.com

Min Soo Kim
minsoo07.kim@samsung.com

A. Using the Average of Inverted Importance Maps

We perform additional experiments to confirm that EVET outperforms the case of simply using the average of inverted importance maps. We consider average drop, increase in confidence, and the energy-based pointing game on VOC segmentation test set (210 images). The results are given in Table 1. As can be seen, EVET achieves the best performance in all cases, which shows the superiority of EVET. On the other hand, using the average makes little improvement and even has worse performance than the baseline methods except for Gradient.

	Gradient	Grad-CAM	Grad-CAM++	Score-CAM
Average drop (%)	24.8 / 18.8 / **16.6**	11.0 / 12.7 / **7.8**	9.0 / 10.5 / **7.0**	3.4 / 4.7 / **2.0**
Increase in confidence (%)	29.5 / 35.2 / **36.7**	50.0 / 46.2 / **55.2**	41.4 / 40.95 / **47.1**	53.3 / 50.5 / **60.0**
Energy-based pointing game (%)	28.0 / 29.7 / **30.9**	47.5 / 47.1 / **50.0**	35.5 / 35.7 / **39.9**	34.8 / 34.9 / **38.4**

Table 1: Comparative evaluation of EVET with the case of using the average of importance maps obtained from the transformed images. Each cell has values of (original / average / EVET).

B. Choice of Transformations for EVET

We consider a set of geometric image transformations: scale, rotate, shear, and horizontal flip since they are widely used in image data augmentation and easy to apply. Parameters of transformations are given in Table 2 and examples are shown in Fig. 1. To investigate how the choice of transformations and their parameters affects the performance of EVET, we compute average drop, increase in confidence, win, and the energy-based pointing game on VOC segmentation test set (210 images) when a subset of transformations is used for EVET (using $\alpha = 0$). We then rank each subset based on the score defined by:

$$\text{Total score} = -AD + IC + Win + EBP \tag{1}$$

where AD, IC, Win, and EBP represent the normalized average drop, increase in confidence, win, and energy-based pointing game, respectively. In the above, $-AD$ is used since lower is better for average drop; higher is better for the others. The results are given in Table 3 and 4. In addition, we calculate the difference between the original map and the inverted map obtained from the transformed image to see what transformations result in more variance (column dissimilarity in the tables). For $n = 1, 2, 3$, $S_{4,n}$ represents a set of transformations which consists of horizontal flip, scale by $1 - 0.1n$, rotate by $10n°$, and shear by $0.2n$; $S_{7,n}$ consists of the transformations in $S_{4,n}$ and the transformations that apply these scale, rotate, shear after horizontal flip.

We report several common observations for Gradient and Grad-CAM:

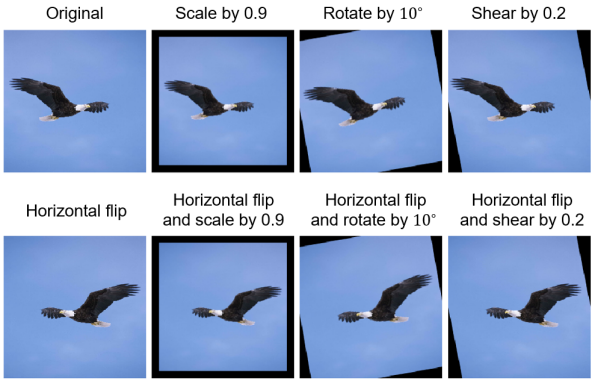


Figure 1: Examples of transformed images used for EVET in our experiments.

Transformation	Parameters
Scale	$0.9, 0.8, 0.7$
Rotation	$10°, 20°, 30°$
Shear	$0.1, 0.2, 0.3$

Table 2: Transformation parameters used for EVET.

- Larger difference between the original and transformed images makes larger dissimilarity between the original and resulting importance maps when using a single transformation. For example, dissimilarity increases as the rotation angle increases. Similarly, applying the transformation after horizontal flip leads to larger dissimilarity than when the transformation is solely applied.
- EVET tends to show better performance when more transformations are used. It is readily seen that $S_{4,n}$'s have better performance than using a single transformation and $S_{7,n}$'s have better performance than $S_{4,n}$'s. Therefore, we decide to use $S_{7,1}$ for our experiments.

We report different patterns in Gradient and Grad-CAM as follows:

- For Gradient, scale achieves better performance than rotate. On the other hand, horizontal flip shows the worst performance.
- For Grad-CAM, horizontal flip shows high performance among the cases of using a single transformation. Furthermore, combination of horizontal flip and other transformations has a positive effect. We observe that applying rotate and shear after horizontal flip improves the performance.

Transformation	Average drop (%)	Increase in confidence (%)	Win (%)	Energy-based pointing game (%)	Dissimilarity	Total score	Rank
Horiznotal flip	20.59	30.95	58.1	28.2	10.95	-5.2	25
Scale (0.9)	19.45	36.67	70	29.06	13.27	0.52	11
Scale (0.8)	16.98	37.62	68.1	29.15	14.94	1.61	9
Scale (0.7)	15.39	37.62	64.29	28.71	18.58	0.63	10
Flip & scale (0.9)	18.24	35.71	66.67	29.14	13.6	0.39	12
Flip & scale (0.8)	16.29	40	69.52	29.12	15.04	2.67	7
Flip & scale (0.7)	15	40.95	67.62	28.8	18.83	2.39	8
Rotate (10)	18.49	35.71	67.14	28.08	12.76	-2.07	15
Rotate (20)	18.03	35.24	67.62	28.17	13.62	-1.75	13
Rotate (30)	17.98	34.29	61.9	28.03	14.29	-3.16	21
Flip & rotate (10)	18.55	34.76	62.86	28.03	13.31	-3.09	20
Flip & rotate (20)	18.17	35.71	61.9	28.35	13.92	-2.1	16
Flip & rotate (30)	17.68	36.67	62.38	28.09	14.65	-2.18	17
Shear (0.9)	19.94	32.86	65.71	28.03	11.27	-3.71	22
Shear (0.8)	17.55	36.67	64.29	28.06	12.5	-1.92	14
Shear (0.7)	17.99	33.33	64.76	28.29	13.84	-2.41	18
Flip & shear (0.9)	20.23	32.86	58.57	28.21	12.14	-4.45	24
Flip & shear (0.8)	19.69	33.81	65.24	28.26	13.22	-2.9	19
Flip & shear (0.7)	20.4	32.38	61.43	28.23	14.45	-4.18	23
$S_{4,1}$	14.95	40.95	76.67	28.64	N/A	3.36	6
$S_{4,2}$	14.63	41.43	77.62	28.8	N/A	4.12	5
$S_{4,3}$	12.4	40.95	75.24	28.67	N/A	4.16	4
$S_{7,1}$	13.51	40.95	80	28.98	N/A	5.16	3
$S_{7,2}$	12.81	42.38	81.9	29.21	N/A	6.62	2
$S_{7,3}$	10.4	44.29	80	29.1	N/A	7.5	1

Table 3: Performance of EVET with different subsets of transformations for Gradient on VOC.

Transformation	Average drop (%)	Increase in confidence (%)	Win (%)	Energy-based pointing game (%)	Dissimilarity	Total score	Rank
Horiznotal flip	9.55	51.9	53.81	47.39	8.05	0.5	9
Scale (0.9)	9.39	50.95	47.14	47.21	9.87	-1.09	17
Scale (0.8)	8.26	51.9	45.71	46.37	12.88	-0.93	16
Scale (0.7)	8.05	51.9	45.24	45.13	17.84	-2.61	22
Flip & scale (0.9)	8.76	50.48	47.62	47.31	11.01	-0.55	14
Flip & scale (0.8)	8.65	50.95	45.71	46.69	13.07	-1.36	18
Flip & scale (0.7)	8.7	50.48	43.81	45.35	17.49	-3.9	24
Rotate (10)	9.31	50	44.76	47.43	10.2	-1.57	19
Rotate (20)	8.76	51.9	42.86	47.21	12.55	-0.58	15
Rotate (30)	9.58	50.48	40.48	46.97	14.58	-2.83	23
Flip & rotate (10)	8.98	51.9	47.14	47.42	11.34	0.13	13
Flip & rotate (20)	9.19	53.33	49.52	47.27	13.21	0.86	7
Flip & rotate (30)	8.71	53.81	43.81	46.88	15.16	0.2	12
Shear (0.9)	10.37	50.48	47.62	47.25	8.45	-2.15	20
Shear (0.8)	9.48	51.43	42.38	46.8	11.19	-2.18	21
Shear (0.7)	10.63	50.48	37.62	46.66	14.21	-4.68	25
Flip & shear (0.9)	9.5	51.9	53.33	47.24	10.35	0.26	11
Flip & shear (0.8)	9.34	53.33	50	47.22	12.06	0.71	8
Flip & shear (0.7)	9.1	54.76	42.86	47.06	14.63	0.49	10
$S_{4,1}$	8.3	53.81	59.52	47.13	N/A	3.18	5
$S_{4,2}$	7.03	54.76	59.05	46.43	N/A	3.82	4
$S_{4,3}$	7.54	53.33	55.24	45.69	N/A	0.91	6
$S_{7,1}$	7.56	54.29	62.38	47.14	N/A	4.56	1
$S_{7,2}$	6.74	54.76	61.9	46.44	N/A	4.51	2
$S_{7,3}$	6.21	56.67	59.52	45.44	NaN	4.28	3

Table 4: Performance of EVET with different subsets of transformations for Grad-CAM on VOC.

C. Hyper-parameter α

Similar to the previous section, we consider average drop, increase in confidence, win, and the energy-based pointing game on VOC and COCO to investigate the effect of the choice of α. We choose the best value of α based on the score defined in the previous section. The results are provided in Table 5 and 6.

In general, increasing α encourages the resulting saliency map to be more concentrated on the target object. It turns out that increasing α improves performance in terms of the pointing game while the target class probabilities decrease. In case of ImageNet, we use $\alpha = 0.1$ which shows high performance overall.

	AD	IC	Win	EBP	Rank
Original	24.8	29.5	0	28	6
$\alpha = \mathbf{0.1}$	16.6	36.7	70.5	30.9	1
$\alpha = 0.2$	21	35.2	59	32.6	2
$\alpha = 0.3$	24.8	32.9	53.3	33.9	3
$\alpha = 0.4$	29.6	31	48.6	34.9	4
$\alpha = 0.5$	32.6	30	44.8	35.6	5

(a) Gradient

	AD	IC	Win	EBP	Rank
Original	11	50	0	47.5	6
$\alpha = \mathbf{0.1}$	7.8	55.2	51.9	50	1
$\alpha = 0.2$	8.6	51	46.7	51.5	2
$\alpha = 0.3$	9.7	50	44.3	52.6	3
$\alpha = 0.4$	10.3	49.5	42.4	53.4	4
$\alpha = 0.5$	11	48.1	35.7	54	5

(b) Grad-CAM

	AD	IC	Win	EBP	Rank
Original	9	41.4	0	35.5	6
$\alpha = 0.1$	6.7	47.1	59	38.1	2
$\alpha = \mathbf{0.2}$	7	47.1	55.2	39.9	1
$\alpha = 0.3$	7.7	44.3	49.5	41.2	3
$\alpha = 0.4$	8.2	43.8	46.2	42.2	4
$\alpha = 0.5$	9	43.8	43.8	43	5

(c) Grad-CAM++

	AD	IC	Win	EBP	Rank
Original	3.4	53.3	0	34.8	6
$\alpha = 0.1$	1.8	58.1	57.6	37	2
$\alpha = \mathbf{0.2}$	2	60	54.8	38.4	1
$\alpha = 0.3$	2.4	58.1	50.5	39.5	3
$\alpha = 0.4$	2.6	58.1	50.5	40.3	4
$\alpha = 0.5$	2.9	57.6	49	41	5

(d) Score-CAM

Table 5: Choice of α for EVET on VOC. AD, IC, Win, and EBP represent average drop, increase in confidence, win, and the energy-based pointing game, respectively.

	AD	IC	Win	EBP	Rank
Original	24.4	40.1	0	27.4	6
$\alpha = \mathbf{0.1}$	17.4	47.5	69.5	30.5	1
$\alpha = 0.2$	20.2	44.7	60.4	32.4	2
$\alpha = 0.3$	23.2	41.3	52.5	33.8	3
$\alpha = 0.4$	26.3	36.7	47.7	35	4
$\alpha = 0.5$	29.4	34.8	43	35.9	5

(a) Gradient

	AD	IC	Win	EBP	Rank
Original	15.3	57.5	0	44.9	6
$\alpha = \mathbf{0.1}$	12.2	60.4	56	47.2	1
$\alpha = 0.2$	13.9	58.1	47.3	48.7	2
$\alpha = 0.3$	15.2	55.7	41.8	49.8	3
$\alpha = 0.4$	16.4	53	38.1	50.6	4
$\alpha = 0.5$	17.3	51.1	35.8	51.3	5

(b) Grad-CAM

	AD	IC	Win	EBP	Rank
Original	13.2	52.4	0	33.6	6
$\alpha = \mathbf{0.1}$	10.7	55.4	61.3	36.2	1
$\alpha = 0.2$	11.4	54	53.9	37.9	2
$\alpha = 0.3$	12	52.4	49.1	39.1	3
$\alpha = 0.4$	12.6	51.5	46.3	40.1	4
$\alpha = 0.5$	13.3	49.5	43.2	40.8	5

(c) Grad-CAM++

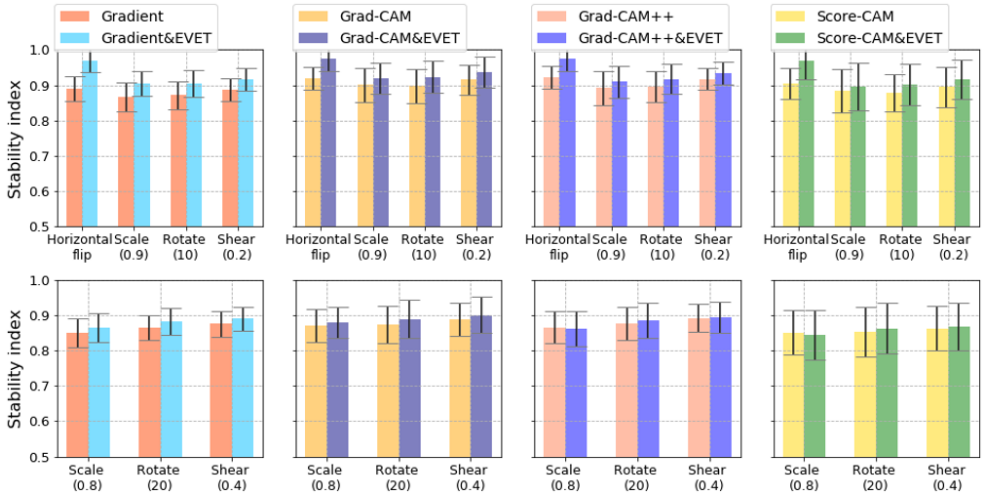
	AD	IC	Win	EBP	Rank
Original	7.7	59.8	0	31.7	6
$\alpha = \mathbf{0.1}$	5.3	64.6	66.6	33.8	1
$\alpha = 0.2$	5.6	63.9	61	35	2
$\alpha = 0.3$	6	62.2	56.5	36	3
$\alpha = 0.4$	6.4	60.3	53.3	36.7	4
$\alpha = 0.5$	6.9	58.7	51.5	37.3	5

(d) Score-CAM

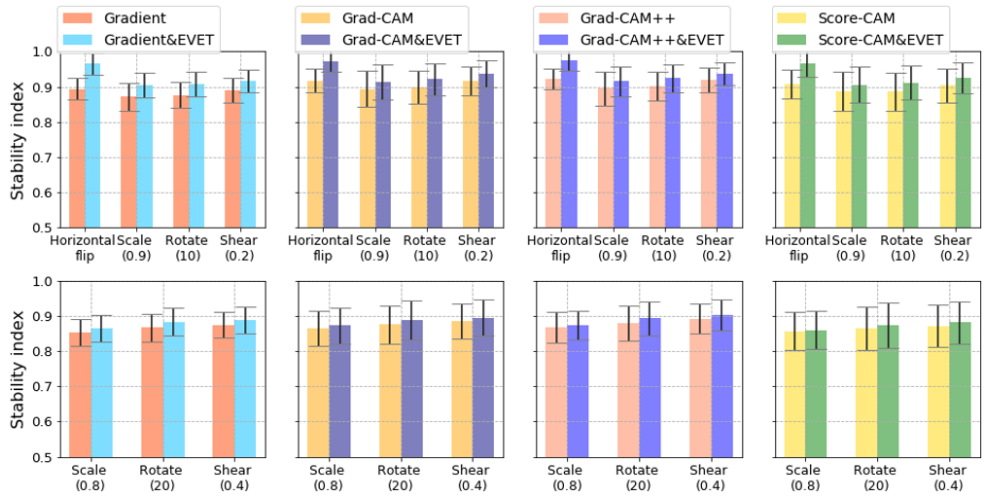
Table 6: Choice of α for EVET on COCO. AD, IC, Win, and EBP represent average drop, increase in confidence, win, and the energy-based pointing game, respectively.

D. Stability Evaluation with Respect to Image Transformations on VOC and COCO

Fig. 2 shows that applying EVET leads to an increase in SI.



(a) Stability evaluation of EVET on VOC.



(b) Stability evaluation of EVET on COCO

Figure 2: Stability evaluation with respect to image transformations on VOC and COCO.