Manitest: Are classifiers really invariant?

Alhussein Fawzi
alhussein.fawzi@epfl.ch
Pascal Frossard
pascal.frossard@epfl.ch

Signal Processing Laboratory (LTS4)
Ecole Polytechnique Fédérale de Lausanne (EPFL)
Lausanne, Switzerland

Invariance to geometric transformations is a highly desirable property of classifiers in image recognition tasks. Nevertheless, it is unclear to which extent state-of-the-art classifiers are invariant to transformations such as rotations and translations. This is mainly due to the lack of general methods that properly measure such an invariance. We propose *Manitest*, a rigorous and systematic approach for quantifying the invariance to geometric transformations of an arbitrary classifier. The source code of Manitest is available at the project website `http://sites.google.com/site/invmanitest/`.

Definition of *Manitest* measure. For a fixed image I, we measure the robustness of a classifier f relatively to the transformation group $\mathcal{T}$ as the minimal normalized *distance* between the identity transformation e and a transformation $\tau \in \mathcal{T}$ that changes the classification label when applied to the image.

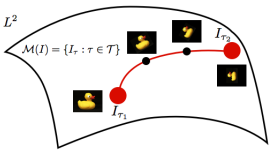
$$\Delta_{\mathcal{T}}(I;f) = \min_{\tau \in \mathcal{T}} \frac{d(e,\tau)}{\|I\|_{L^2}} \text{ subject to } f(I_\tau) \neq f(I),$$

where
$\mathcal{T}$: transformation group,
I_τ : I transformed by τ,
d : distance on $\mathcal{T}$,
e : identity transformation.

For a distribution of samples μ, the global invariance measure of f is obtained by averaging $\Delta_{\mathcal{T}}(I;f)$

$$\rho_{\mathcal{T}}(f) = \mathbb{E}_{I\sim\mu}\Delta_{\mathcal{T}}(I;f).$$

Which distance to use on $\mathcal{T}$? A crucial element in the definition of our Manitest invariance measure $\rho_{\mathcal{T}}$ is the choice of the distance measure d. In order to define d, our novel key idea is to represent the set of transformed versions of an image as a manifold; the transformation metric is naturally captured by the geodesic distance on the manifold. For a given image, the invariance measure $\Delta_{\mathcal{T}}(I;f)$ therefore corresponds to the minimal normalized *geodesic* distance on the manifold that leads to a point where the classifier's decision is changed.



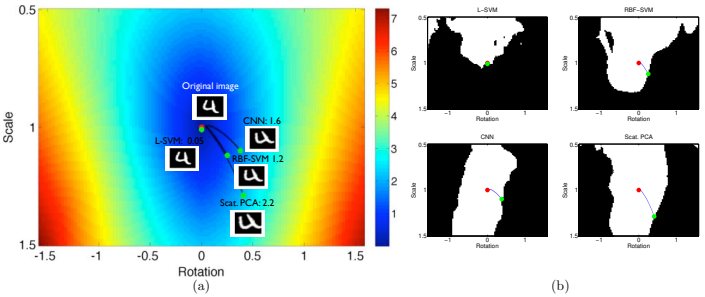
Algorithm for computing $\Delta_{\mathcal{T}}(I;f)$. The manifold $\mathcal{T}$ is sampled using a regular grid, and the geodesic distances are estimated using the *Fast Marching* algorithm [2, 3]. The algorithm is terminated whenever a node that changes the classifier's decision is visited.

Experimental results. A comparison of different classifiers in terms of their Manitest invariance scores on the MNIST digit classification dataset is shown in Table 1. The classifier based on scattering transforms outperforms other classifiers in terms of robustness to translations and global similarity transformations. We visualize the invariance scores of the different tested classifiers on an example test image shown in Fig. 1.

Group	Lin. SVM	RBF-SVM	CNN	Scat. [1]
Test error (%)	8.4	1.4	**0.7**	0.8
Translations	0.8	1.3	1.7	**2.1**
Scale + Rotation	0.8	1.5	**1.9**	1.8
Similarity	0.6	1.1	1.5	**1.6**

Table 1: Accuracy and invariance scores of different classifers on the MNIST dataset.

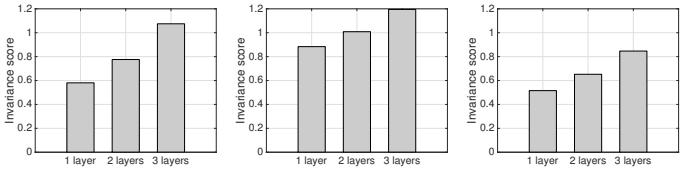
Using Manitest, we also quantify the effect of *data augmentation* and *CNN depth* on the invariance of a classifier. Our result shows that the invariance score of an RBF-SVM increases by 50% on the similarity group



(a) (b)

Figure 1: Visualization of invariance to the Scale+Rotation group for an example image of digit "4". **(a):** Distance map, where the color code indicates the geodesic distance from the identity transformation (shown by red dot at the center). For each classifier, the minimal transformation for which the output of the classifier is not correct (i.e., not "4") is indicated, along with the corresponding transformed image and geodesic path. **(b):** The region where the classifier correctly outputs the label "4" is shown in white. Geodesic paths are also shown.

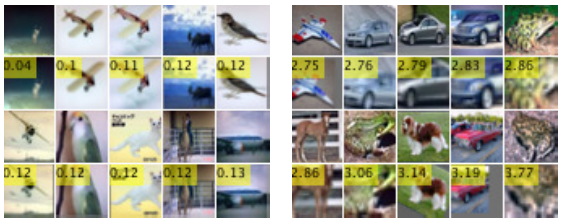
by merely adding transformed samples in the training set. Quite surprisingly, the invariance score of the RBF-SVM trained with augmented samples surpasses that of a CNN (without augmentation), which shows the merits of data augmentation in terms of increasing invariance. Moreover, we show in Fig. 2 the increasing invariance of Manitest scores with the number of layers of a CNN on the CIFAR-10 dataset.



(a) Translations (b) Dilation + Rotation (c) Similarity

Figure 2: Invariance scores of CNNs on $\mathcal{T}_{\text{trans}}$, $\mathcal{T}_{\text{dil+rot}}$ and $\mathcal{T}_{\text{sim}}$, for the CIFAR-10 dataset.

Fig. 3 shows a ranking of the images in terms of their invariance scores $\Delta_{\mathcal{T}}(I;f)$. Despite the high accuracy of the three layer CNN on the CIFAR-10 task, note that a slight transformation of the original images can change the classification label of the classifier (see worst 10 images). More emphasis should therefore be put in order to achieve higher levels of invariance.



(a) Worst 10 (b) Top 10

Figure 3: Illustration of images having (a) worst, (b) top invariance to similarity transformations, for the three-layer CNN. The odd rows show the original images, and the even rows the minimally transformed images changing the prediction of the CNN. The Manitest invariance score is indicated on each transformed image.

[1] Joan Bruna and Stéphane Mallat. Invariant scattering convolution networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(8):1872–1886, 2013.

[2] Ron Kimmel and James A Sethian. Computing geodesic paths on manifolds. *Proceedings of the National Academy of Sciences*, 95(15):8431–8435, 1998.

[3] John N Tsitsiklis. Efficient algorithms for globally optimal trajectories. *IEEE Transactions on Automatic Control*, 40(9):1528–1538, 1995.