

---

# Data augmentation with Möbius transformations

---

Sharon Zhou<sup>1</sup> Jiequan Zhang<sup>1</sup> Hang Jiang<sup>1</sup> Torbjörn Lundh<sup>2</sup> Andrew Y. Ng<sup>1</sup>

## Abstract

Data augmentation has led to substantial improvements in the performance and generalization of deep models, and remain a highly adaptable method to evolving model architectures and varying amounts of data—in particular, extremely scarce amounts of available training data. In this paper, we present a novel method of applying Möbius transformations to augment input images during training. Möbius transformations are bijective conformal maps that generalize image translation to operate over complex inversion in pixel space. As a result, Möbius transformations can operate on the sample level and preserve data labels. We show that the inclusion of Möbius transformations during training enables improved generalization over prior sample-level data augmentation techniques such as cutout and standard crop-and-flip transformations, most notably in low data regimes.

## 1. Introduction

Data augmentation has significantly improved the generalization of deep neural networks on a variety of image tasks, including image classification (Perez & Wang, 2017; DeVries & Taylor, 2017), object detection (Zhong et al., 2017; Guo & Gould, 2015), and instance segmentation (Wang et al., 2018). Prior work has shown that data augmentation on its own can perform better than, or on par with, highly regularized models using other regularization techniques such as dropout (Hernández-García & König, 2018). This effectiveness is especially prominent in low data regimes, where models often fail to capture the full variance of the data in the training set (Zhang et al., 2019).

Many data augmentation techniques rely on priors that are present in the natural world. Standard operations, such as translation, crop, and rotation, in addition to more recent

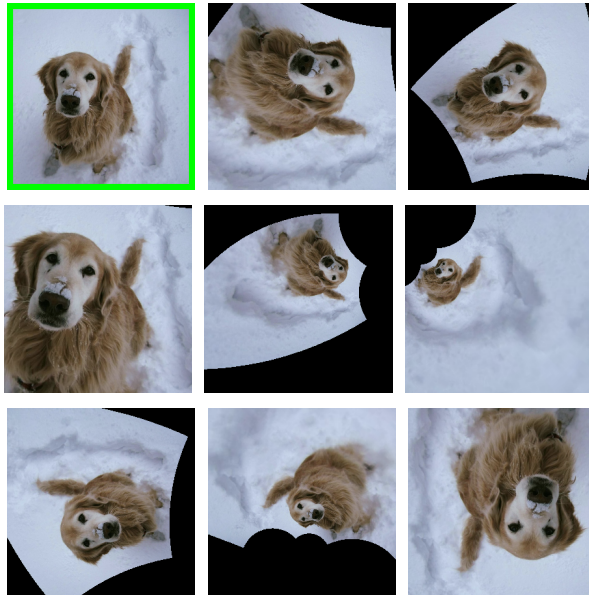


Figure 1. Examples of Möbius transformations on an image (original with green border). The transformations result in variations in perspective, orientation, and scale, while still preserving local angles and anharmonic ratios.

methods, such as cutout (DeVries & Taylor, 2017), have improved generalization by encouraging equivariance to the transformation. For example, an image of a horse remains a horse in its vertical reflection or with its body partially occluded. As a result, these transformations are able to preserve the original label on the augmented sample, enabling straightforward and easy incorporation into the growing number of data augmentation algorithms for both fully supervised and semi-supervised learning. In a nutshell, these sample-level methods have been not only effective, but also interpretable, easy to implement, and flexible to incorporate.

Following the success of these methods, we focus this paper on augmentations that exploit natural patterns to preserve labels after transformation and that operate on the sample level. These transformations easily complement other methods, thus leveraged in a wide variety of data augmentation algorithms. In contrast, multi-sample augmentations, which have had comparably strong empirical results (Zhang et al., 2017), unfortunately connect less clearly to natural priors

---

<sup>1</sup>Department of Computer Science, Stanford University, Stanford, United States <sup>2</sup>Department of Mathematics, Chalmers University of Technology, Gothenberg, Sweden. Correspondence to: Sharon Zhou <sharonz@cs.stanford.edu>.

that would support equivariance to the augmentation. While performant on their own, these methods have had less success with integration into data augmentation algorithms and policies (Cubuk et al., 2019a;b; Xie et al., 2019), except for those tailored to them (Berthelot et al., 2019).

In this paper, we propose a novel data augmentation technique, inspired by biological patterns, using bijective conformal maps known as *Möbius transformations*. Möbius transformations perform complex inversion in pixel space, extending standard translation to include divisibility. These transformations enable perspective projection—or transforming perceived distance of objects in an image—and are found naturally in the anatomy and biology of humans and other animals.

We apply Möbius transformations as a form of data augmentation during training to improve the generalization of deep models. We show empirically that the inclusion of Möbius transformations improves performance on the CIFAR-10 and CIFAR-100 benchmarks over prior sample-level data augmentation techniques, such as cutout (DeVries & Taylor, 2017) and standard crop-and-flip baselines. Möbius transformations are especially useful in low data settings particularly when applied in small quantities. Finally, we show that Möbius can complement other transformations, including cutout and crop-and-flip.

Our key contributions can be summarized as follows:

- **Method:** We introduce Möbius transformation as a method for data augmentation on images. This method works well at a small amount of inclusion, i.e. 20% of training set or only 80 images per class in the reduced CIFAR-10 dataset.
- **Performance:** Empirically, the inclusion of Möbius data augmentation improves neural network generalization over prior methods that use sample-level augmentation techniques, such as cutout (DeVries & Taylor, 2017) and standard crop-and-flip transformations.
- **Low data:** Möbius transformations are especially effective in low data settings, where the data quantity is on the order of hundreds of samples per class.
- **Complementary use:** Möbius data augmentation is designed to complement other methods, e.g. cutout. Implemented together, they can outperform either alone.

## 2. Möbius transformations

Möbius transformations are bijective conformal mappings that operate over complex inversion and preserve local angles. They are also known as bilinear or linear fractional transformations. We discuss their biological and perceptual

underpinnings, and follow with a formal definition. Finally, we describe their application to data augmentation to improve generalization in convolutional neural networks.

### 2.1. Motivation

Möbius transformations have been studied in biology as 2D projections of specimens—such as humans, fungi, and fish—from their 3D configurations (Thompson et al., 1942; Petukhov, 1989; Lundh et al., 2011). Mathematically, most of these examples leverage Liouville’s theorem (Liouville, 1850), which states that smooth conformal mappings are Möbius transformations on a domain of  $\mathbb{R}^n$  where  $n > 2$ . These biological patterns motivate our application of Möbius transformations to natural images, particularly those that include the relevant species.

Beyond biological underpinnings, Möbius transformations preserve the *anharmonic ratio* (Ahlfors, 1989; Needham, 1998), or the extent to which four collinear points on a projective line deviate from the harmonic ratio.<sup>1</sup> This invariance is a property that Möbius transformations share with projective transformations, which are used widely in metrology (Criminisi et al., 2000). In the context of transforming natural images, such a transformation can be particularly useful for perspective projection. That is, an image can be transformed to an alternate perceived distance. This effect is visually apparent across examples in Figure 1.

### 2.2. Definition

Existing data augmentation techniques for image data belong to the class of affine mappings, i.e. the group of translation, scaling, and rotation, which can be generally described using a complex function  $z \rightarrow az + b$ , where the variable  $z$  and the two parameters  $a$  and  $b$  are complex numbers. Möbius transformations represent the next level of abstraction by introducing division to the operation (Lie, 1893; Petukhov, 1989). The group of Möbius transformations can be described as all functions  $f$  from  $\mathbb{C} \rightarrow \mathbb{C}$  with the form

$$f(z) = \frac{az + b}{cz + d}, \quad (1)$$

where  $a, b, c, d \in \mathbb{C}$  such that  $ad - bc \neq 0$ . As a result, the set of all Möbius transformations is a superset of several basic transformations, including translation, rotation, inversion, and even number of reflections over lines.

One method for programatically specifying the different parameters  $a, b, c, d$  in Eq. (1) is to use the fact that there exists a unique Möbius transformation sending any three points to any three other points in the extended complex plane (Needham, 1998, p. 150). That is, instead of specifying the four

<sup>1</sup>The anharmonic ratio, also denoted *cross-ratio*, stems from projective geometry and has been studied in biology with respect to Möbius transformations (Petukhov, 1989; Lundh et al., 2011).

complex parameters  $a, b, c$  and  $d$  directly in (1), we can define three separate points  $z_1, z_2, z_3 \in \mathbb{C}$  in the image and then select three separate target points  $w_1, w_2, w_3 \in \mathbb{C}$ , to which those initial points will be mapped in the resulting transformation. From these two sets of points, we can then compute the parameters of the transformation using the knowledge that anharmonic ratios—adding the points  $z_i$  and  $w_i$  where  $i = \{1, 2, 3\}$  completes the two quartets—are Möbius invariant (Needham, 1998, p. 154), resulting in the following equality:

$$\frac{(w - w_1)(w_2 - w_3)}{(w - w_3)(w_2 - w_1)} = \frac{(z - z_1)(z_2 - z_3)}{(z - z_3)(z_2 - z_1)}. \quad (2)$$

We can rearrange this expression by solving for  $w$ :

$$\frac{w - w_1}{w - w_3} = \frac{(z - z_1)(z_2 - z_3)(w_2 - w_1)}{(z - z_3)(z_2 - z_1)(w_2 - w_3)}$$

$$w = \frac{Aw_3 - w_1}{A - 1}$$

where  $A = \frac{(z - z_1)(z_2 - z_3)(w_2 - w_1)}{(z - z_3)(z_2 - z_1)(w_2 - w_3)}$ . This final expression for  $w$  is in the form of Eq. (1):

$$f(z) = w = \frac{Aw_3 - w_1}{A - 1} = \frac{az + b}{cz + d},$$

from which we can compute the following values for  $a, b, c$ , and  $d$  using basic algebraic operations:

$$a = w_1w_2z_1 - w_1w_3z_1 - w_1w_2z_2$$

$$+ w_2w_3z_2 + w_1w_3z_3 - w_2w_3z_3,$$

$$b = w_1w_3z_1z_2 - w_2w_3z_1z_2 - w_1w_2z_1z_3$$

$$+ w_2w_3z_1z_3 + w_1w_2z_2z_3 - w_1w_3z_2z_3,$$

$$c = w_2z_1 - w_3z_1 - w_1z_2 + w_3z_2 + w_1z_3 - w_2z_3,$$

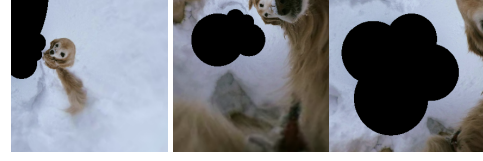
$$d = w_1z_1z_2 - w_2z_1z_2 - w_1z_1z_3$$

$$+ w_3z_1z_3 + w_2z_2z_3 - w_3z_2z_3.$$

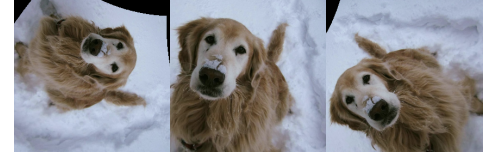
Alternatively, by solving Eq. (2) using linear algebra, i.e. evaluating a determinant from this construction using the Laplace expansion, one can elegantly express these algebraic expressions above as determinants:

$$a = \begin{vmatrix} z_1w_1 & w_1 & 1 \\ z_2w_2 & w_2 & 1 \\ z_3w_3 & w_3 & 1 \end{vmatrix}, \quad b = \begin{vmatrix} z_1w_1 & z_1 & w_1 \\ z_2w_2 & z_2 & w_2 \\ z_3w_3 & z_3 & w_3 \end{vmatrix},$$

$$c = \begin{vmatrix} z_1 & w_1 & 1 \\ z_2 & w_2 & 1 \\ z_3 & w_3 & 1 \end{vmatrix}, \quad d = \begin{vmatrix} z_1w_1 & z_1 & 1 \\ z_2w_2 & z_2 & 1 \\ z_3w_3 & z_3 & 1 \end{vmatrix}.$$



(a) Random Möbius parameterization



(b) Defined Möbius parameterization

Figure 2. Examples of Möbius transformations with fully random parameters that may not preserve labels in 2(a), contrasted with those in our Möbius data augmentation set with defined parameters in 2(b). Note that our method is not devoid of randomness: at a training step  $i$ , a parameterization in the defined set  $f_j \in F$  is randomly selected and applied to the candidate image  $f_j(x_i)$ .

This pointwise method is used in our work to construct valid image augmentations using Möbius transformations. Ultimately, this method can be leveraged to define parameters for specific types of Möbius transformations programmatically for needs within and beyond data augmentation.

### 2.2.1. EQUIVALENT FRAMING: CIRCLE REFLECTION

We introduce an equivalent formulation of Möbius transformations on images in  $\mathbb{R}^2$ . The goal of this section is to lend intuition on constraints that we apply to Möbius data augmentation in Section 2.3 that follows.

Möbius mappings in the plane can also be defined as the set of transformations with an even number of reflections over circles and lines (i.e. circles with infinite radii) on the plane. A reflection, or inversion, in the unit circle is the complex transformation (Needham, 1998, pp. 124):

$$z \in \mathbb{C} \rightarrow \frac{z}{|z|^2}.$$

Thus, a Möbius transformation on an image is simply a reflection over the unit circle, with pixels inside of the circle projected outwards and pixels on the outside projected inwards. As such, Möbius transformations often reflect a different amount of pixels inwards as opposed to outwards, and this imbalance enables the scale distortions seen in Figure 1. Note that a circular shape can be left as an artifact after the transformation, e.g. if the reflection occurs at an edge without any pixels to project inwards (see Figure 2).

### CIFAR-10 images under Möbius transformations

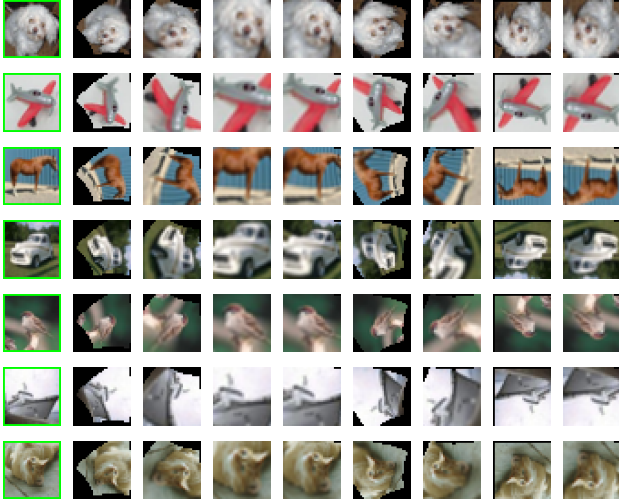


Figure 3. Examples of CIFAR classes {dog, airplane, horse, automobile, bird, ship, cat} undergoing different Möbius transformations with defined parameters (original images have green borders).

### 2.3. Data augmentation

We incorporate Möbius transformations into classical label-preserving data augmentation methods of the form  $(x, y) = (f(x), y)$ , where  $f$  is a Möbius transformation on an image  $x$ , preserving label  $y$ .

In order to use Möbius transformations for data augmentation, we need to constrain the set of possible transformations. When the parameters are taken to the limit, transformations do not necessarily preserve the image label. Note that this is similar to constraining translation in order to ensure that pixels remain afterwards, or to keeping cutout to lengths judiciously less than the size of the image so that it is not fully occluded. Because Möbius transformations inherently reflect more pixels in one direction (into or out of the circle), we will often see two main effects: (1) incongruent sizes of the output from the initial input and (2) gaps between pixels in the result transformation, sometimes significant depending on the location of the circle. E.g., if the circle is placed at the edge of the image, there is little to project from the edge inwards. To address both of these effects, we enforce equal sizing after the transformation and cubic spline interpolation during reflection to fill gaps.

Provided that the group of Möbius transformations can result in reflections with arbitrarily large scale distortions, we define a set of Möbius transformations for data augmentation. These eight Möbius transformations preserve labels with limited gaps. We juxtapose Möbius transformations defined in our set with those that are randomly parameterized in Figure 2 for visual comparison.

The Möbius transformations in our defined set modify the

two sets of three points  $z_i$  and  $w_i$  where  $i = \{1, 2, 3\}$  that perform the mapping. The exact parameters are complex functions of image dimensions (detailed in Appendix A). Examples of these are shown on different CIFAR-10 classes in Figure 3. Based on their visual appearance, we describe them each as follows: (1) clockwise twist, (2) clockwise half-twist, (3) spread, (4) spread twist, (5) counter clockwise twist, (6) counter clockwise half-twist, (7) inverse, and (8) inverse spread.

For a given step in training, a Möbius transformation in the defined set is randomly selected and applied to the image. This process is similar to RandAugment (Cubuk et al., 2019b), a data augmentation policy that randomly selects among a constrained space of possible augmentations. We note that, similar to RandAugment, our defined variations are not the limit to useful Möbius transformations on image data; they demonstrate the first step to effectively incorporating them into data augmentation.

## 3. Related work

A large number of data augmentation techniques have recently emerged for effectively regularizing neural networks, including both sample-level augmentations, such as ours, as well as multi-sample augmentations that mix multiple images. We discuss these, as well as data augmentation algorithms that leverage multiple augmentations. Finally, we examine ways in which Möbius transformations have been applied to deep learning. To our knowledge, this is the first work using Möbius transformations for data augmentation in deep neural networks.

### 3.1. Sample-level augmentation

Möbius transformations generalize standard translation to include inversion as an operation under conformity, demonstrating outputs that appear to have gone through crop, rotation, and/or scaling, while preserving local angles from the original image. We recognize that the list of image transformations is extensive: crop, rotation, warp, skew, shear, random distortion, Gaussian blur, Gaussian noise, among many others. Additional sample-level data augmentation methods use occlusion such as cutout (DeVries & Taylor, 2017) and random erasing (Zhong et al., 2017), which apply random binary masks across image regions.

### 3.2. Multi-sample augmentation

Data augmentation on images also consists of operations applied to multiple input images. In such cases, original labels are often mixed. For example, MixUp (Zhang et al., 2017) performs a weighted average of two images (over pixels) and their corresponding labels in varying proportions to perform soft multi-label classification. Between-class



Augmentation Method	Dataset	Average # Images Per Class	Total # Training Images	Accuracy
Crop-and-flip	CIFAR-10	5000	50k	96.47% $\pm$ 0.04
Cutout <sub>l=16</sub>	CIFAR-10	5000	50k	97.13% $\pm$ 0.03
Möbius	CIFAR-10	5000	50k	96.72% $\pm$ 0.06
Möbius + Cutout <sub>l=16</sub>	CIFAR-10	5000	50k	<b>97.19% <math>\pm</math>0.07</b>
Crop-and-flip	CIFAR-100	600	50k	81.91% $\pm$ 0.20
Cutout <sub>l=8</sub>	CIFAR-100	600	50k	82.35% $\pm$ 0.19
Möbius	CIFAR-100	600	50k	82.85% $\pm$ 0.31
Möbius + Cutout <sub>l=8</sub>	CIFAR-100	600	50k	<b>82.92% <math>\pm</math>0.14</b>
Crop-and-flip	CIFAR-10	400	4k	83.98% $\pm$ 0.16
Cutout <sub>l=16</sub>	CIFAR-10	400	4k	85.20% $\pm$ 0.22
Möbius	CIFAR-10	400	4k	<b>86.07% <math>\pm</math>0.24</b>
Möbius + Cutout <sub>l=16</sub>	CIFAR-10	400	4k	85.70% $\pm$ 0.26

Table 1. Results of main experiments across 5 runs, on several dataset settings: CIFAR-10, CIFAR-100, and reduced CIFAR-10. Möbius with cutout outperforms all other conditions, followed closely by Möbius alone in the CIFAR-100 and reduced CIFAR-10 cases, where the average number of images per class is on the order of hundreds. Next, cutout alone and finally standard crop-and-flip. On CIFAR-10, cutout exceeds Möbius alone, and both outperform the crop-and-flip baseline. We use the optimal lengths for cutout on each dataset (length 16 for CIFAR-10 and length 8 for CIFAR-100) based on prior work (DeVries & Taylor, 2017).

learning (Tokozume et al., 2018) and SamplePairing (Inoue, 2018) are similar techniques, though the latter differs in using a single label. Comparably, RICAP (Takahashi et al., 2019), VH-Mixup and VH-BC+ (Summers & Dinneen, 2019) form composites of several images into one. While these methods have performed well, we focus this paper on comparisons to sample-level augmentations that preserve original labels and that can be more readily incorporated into data augmentation algorithms and policies.

### 3.3. Algorithms and policies for data augmentation

Various strategies have emerged to incorporate multiple data augmentation techniques for improved performance. AutoAugment (Cubuk et al., 2019a), Adatransform (Tang et al., 2019), RandAugment (Cubuk et al., 2019b), and Population Based Augmentation (Ho et al., 2019) offer ways to select optimal transformations (and their intensities) during training. In semi-supervised learning, unsupervised data augmentation, or UDA (Xie et al., 2019), MixMatch (Berthelot et al., 2019), and FixMatch (Sohn et al., 2020) have shown to effectively incorporate unlabeled data by exploiting label preservation and consistency training. Tanda (Ratner et al., 2017) composes sequences of augmentation methods, such as crop followed by cutout then flip, that are tuned to a certain domain. DADA (Zhang et al., 2019) frames data augmentation as an adversarial learning problem and applies this method in low data settings. We do not test all of these augmentation schemes: our results suggest that Möbius transformations could add value as an effective addition to the search space of augmentations, e.g. in AutoAugment, or as a transformation that helps enforce consistency between original and augmented data, e.g. in UDA.

### 3.4. Möbius transformations in deep learning

Möbius transformations have been previously studied across a handful of topics in deep learning. Specifically, they have been used as building blocks in new activation functions (Özdemir et al., 2011) and as operations in hidden layers (Zammit-Mangion et al., 2019). Coupled with the theory of gyrovector spaces, Möbius transformations have inspired hyperbolic neural networks (Ganea et al., 2018). They also play an important component in deep fuzzy neural networks for approximating the Choquet integral (Islam et al., 2019). Finally, model activations and input-output relationships have been theoretically related to Möbius transformations (Mandic, 2000). While prior work has primarily leveraged them for architectural contributions, our work is the first to our knowledge to introduce Möbius transformations for data augmentation and their empirical success on image classification benchmarks.

## 4. Experiments

We experiment on CIFAR-10 and CIFAR-100 image classification benchmarks using standard data splits of 50k training and 10k test (Krizhevsky et al., 2009). CIFAR-10 has 10 classes with 10k images per class, while CIFAR-100 has 100 classes with 500 images per class, in their training sets. We additionally evaluate reduced CIFAR-10, a common low data regime for which only a subset of the training data—4k images with 400 images per class—is used. We do so in a fully supervised manner, without using unlabeled data. Thus, we explore three dataset settings: (1) CIFAR-10, (2) CIFAR-100, and (3) reduced CIFAR-10. The goal of these experiments is to assess the fundamental concept of

### Varying Möbius representation in data augmentation

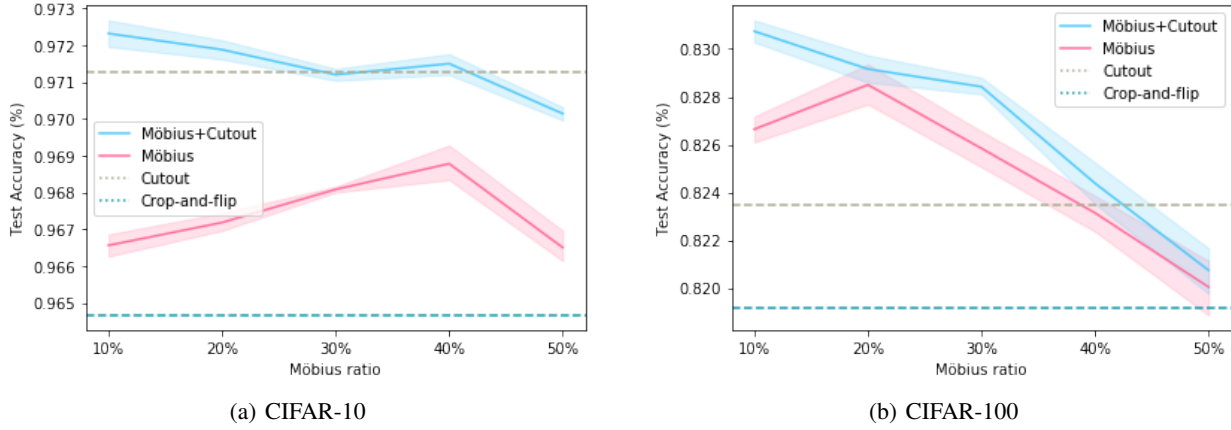


Figure 4. Results from increasing Möbius representation in data augmentation from 10% to 50% in 10% increments, across 5 runs. (a) On CIFAR-10, Möbius at only 10% with cutout demonstrates empirically best results. Möbius on its own performs best at 40%, though it still performs under cutout alone. (b) On CIFAR-100, Möbius reaches best performance at 20% on its own and at 10% with cutout. On both datasets, Möbius boosts the performance of cutout when applied together, particularly in small quantities of 10-30%.

including Möbius data augmentation during training in both regular and low data settings.

#### 4.1. Evaluation of benchmarks

Following prior work on introducing novel data augmentation methods (DeVries & Taylor, 2017; Cubuk et al., 2019a), we use standard crop-and-flip transformations as the baseline across all experimental conditions. We design our experiments to both compare to, and complement, cutout (DeVries & Taylor, 2017), the previous state-of-the-art image transformation that operates on the sample level, preserves labels, and thus has been easy to incorporate into data augmentation policies. Cutout and standard crop-and-flip also remain the default augmentation choices in recent work (Cubuk et al., 2019a). Thus, we compare the following conditions: (1) baseline with only crop and flip (*regular*), (2) cutout, (3) Möbius, and (4) Möbius with cutout. Note that all conditions incorporate crop and flip transformations, following the original cutout paper (DeVries & Taylor, 2017). Because all augmentation techniques are sample-level and preserve labels, they are complementary and can be layered on each other. We further explore these effects by combining Möbius with cutout in our experiments.

Möbius is able to improve generalization when applied to only a small fraction of the data, contrasted with cutout and standard crop-and-flip that are applied to the full dataset. Across all experiments, we apply Möbius transformations to images with 20% probability during training. We perform additional experiments on modulating the representation of Möbius transformations in Section 4.2 where we discuss this insight further. Standard transformations and cutout are applied to all data.

We use the same model and retain the same hyperparameters across all experiments, which we largely draw from prior work on wide residual networks (Zagoruyko & Komodakis, 2016) and cutout (DeVries & Taylor, 2017). Specifically, our model is a wide residual network with a depth of 28, a widening factor of 10, and a dropout probability of 0.3 on all convolutional layers, which we denote WRN-28-10. The WRN-28-10 is trained using SGD with Nesterov momentum (0.9), weight decay ( $5e-4$ ), and a learning rate (0.1) that follows a cosine annealing scheduler. On the full datasets, we evaluate training on 200 epochs and on the reduced, we use 2000 epochs. We measure average performance and standard deviation across 5 runs for every experimental condition. For hardware, we use 4 NVIDIA GeForce GTX 1070 GPUs on all experiments.

On cutout, we use the optimal length values for each dataset (DeVries & Taylor, 2017): a length of 16 on CIFAR-10 ( $\text{cutout}_{l=16}$ ) and a length of 8 on CIFAR-100 ( $\text{cutout}_{l=8}$ ). Möbius is the same across datasets with no fine-tuning.

As shown in Table 1, these experiments highlight several key observations:

- Empirically, Möbius outperforms cutout and standard crop-and-flip on CIFAR-100 and reduced CIFAR-10. On CIFAR-10, cutout performs better than Möbius, though Möbius with cutout outperforms both alone.
- Möbius’s performance gains are especially remarkable, because Möbius is only applied to 20% of the data and still achieves superior performance. This means that only 100 images per class in CIFAR-100 and 80 images per class in reduced CIFAR-10 receive Möbius

Animal Superclass	Crop-&-flip	Möbius ↑	Non-Animal Superclass	Crop-&-flip	Möbius ↑
Aquatic Mammals	73.0%	74.4% ✓	Flowers	86.0%	85.6%
Fish	81.8%	82.4% ✓	Food Containers	79.6%	80.6% ✓
Insects	85.0%	85.8% ✓	Fruit and Vegetables	86.2%	89.6% ✓
Large Omnivores and Herbivores	84.4%	84.8% ✓	Household Electrical Device	86.4%	87.0% ✓
Large Carnivores	83.8%	84.4% ✓	Household Furniture	86.6%	84.8%
Non-insect Invertebrates	81.4%	82.0% ✓	Large Man-made Outdoor Things	89.8%	89.6%
Medium-sized mammals	82.6%	85.2% ✓	Large Natural Outdoor Scenes	85.2%	87.2% ✓
People	64.6%	66.8% ✓	Trees	75.8%	74.4%
Small Mammals	74.8%	79.0% ✓	Vehicles 1	88.6%	90.8% ✓
Reptiles	73.8%	77.2% ✓	Vehicles	91.8%	90.2%

Table 2. Möbius data augmentation consistently improves the accuracy of image classification on animal superclasses (✓ on ↑), as opposed to non-animal superclasses. This empirical observation suggests that Möbius transformations, having been studied in animals, are particularly attuned to improving generalization in these classes. Note that this finding is interesting, though by no means definitive or theoretically potent.

transformations on average.

- Möbius with cutout outperforms both Möbius alone and cutout alone on both full data settings, suggesting that the two techniques are complementary (Möbius without cutout performs best on reduced CIFAR-10). This is important, as we designed Möbius to combine easily with other augmentation methods.

The results of this experiment suggest that a small amount of Möbius data augmentation can improve over cutout, the state-of-the-art performance on sample-level and label-preserving augmentation strategy. This effect is especially prominent in low data regimes, where there are fewer (on the order of hundreds) of samples per class.

#### 4.2. Modulating the inclusion of Möbius

Given the inherent complexity of Möbius transformations, we additionally explore the effects of incorporating an increasing amount of Möbius transformations into the data augmentation process. We evaluate Möbius representations of 10% to 50%, at increments of 10% in between, on CIFAR-10 and CIFAR-100. The goal of this experiment is to examine the effects of modulating Möbius representation during the training process. Note that the experiments in Section 4.1 only focused on a stationary amount (20%) of Möbius.

We compare these increments of Möbius both with and without cutout. We then juxtapose these results with the baseline of cutout alone and that of standard crop-and-flip. We again report average performance and standard deviations across 5 runs on all experimental conditions. The results presented in Figure 4 emphasize the following findings:

- Too much Möbius data augmentation can result in disruptive training and poorer generalization.
- Möbius augmentation nevertheless outperforms both cutout and standard crop-and-flip baselines, across several values of representation particularly at 10% and 20% representation.
- Möbius augmentation alone experiences a local optimum at 40% inclusion on CIFAR-10 and 20% on CIFAR-100.
- Möbius with cutout performs best with a very modest amount (10%) of Möbius. This is expected, as cutout provides additional regularization.

Though not shown in the graph, we also experiment with an even lower representation (5%) of Möbius in the Möbius with cutout condition, in order to observe local optima and a bottoming out effect. We find that 10% still shows superior performance to 5% representation on both datasets. Specifically, Möbius at 5% with cutout performs  $97.18\% \pm 0.14$  on CIFAR-10 and  $82.97\% \pm 0.17$  on CIFAR-100.

#### 4.3. Random parameterization of Möbius

We run experiments comparing the performance of fully randomized parameterization against our defined parameterization on all dataset settings. Recall that our defined set is finite, while there exists infinite possible Möbius transformations in  $\mathbb{R}^2$  under random parameterization. Figure 2 displays a visual comparison.

As shown in Table 3, we observe that random parameterizations of Möbius perform worse than our defined parameterizations, yet still outperforms the baseline of crop-and-flip

Augmentation Method	Dataset	Accuracy
Crop-and-flip	C10	96.47% $\pm$ 0.04
Möbius	C10	<b>96.72% <math>\pm</math> 0.06</b>
Random Möbius	C10	96.54% $\pm$ 0.06
Crop-and-flip	C100	81.91% $\pm$ 0.20
Möbius	C100	<b>82.85% <math>\pm</math> 0.31</b>
Random Möbius	C100	82.30% $\pm$ 0.11
Crop-and-flip	R C10	83.98% $\pm$ 0.16
Möbius	R C10	<b>86.07% <math>\pm</math> 0.24</b>
Random Möbius	R C10	85.58% $\pm$ 0.31

Table 3. Juxtaposition of model performance using randomly parameterized Möbius transformations with those using defined ones. Möbius transformations with random parameters suffers in performance, though still better than the crop-and-flip baseline.

transformations. For parity, we apply Möbius at 20% on both conditions, as performed in the main experiments (Section 4.1).

#### 4.4. Analysis on animal classes

We analyze the predictions from Möbius data augmentation by superclass (Table 2). Specifically, we compare two models trained on CIFAR-100: one with Möbius and with standard crop-and-flip. These 20 superclasses are higher-level aggregations of five classes each (Krizhevsky et al., 2009). In our analysis, we find that the Möbius-trained model improves performance on the 10 animal superclasses {aquatic mammals, fish, insects, large omnivores and herbivores, large carnivores, non-insect invertebrates, medium-sized mammals, people, small mammals, reptiles}. This contrasts with inconsistent performance differences among the non-animal superclasses. Note that Möbius transformations improve overall performance over the standard baseline. These results suggest that Möbius transformations, which have been studied in animals in prior literature (Thompson et al., 1942; Petukhov, 1989; Lundh et al., 2011), are especially effective on these classes in image classification. While this finding is particularly interesting and consistent with Möbius studies in biology, we heed that this observation remains empirical to this study and requires additional examination in order to be conclusive.

#### 4.5. Comparison to data augmentation policies

Comparing empirical results on CIFAR-10 and CIFAR-100, we find that Möbius with cutout performs similarly to two recent state-of-the-art augmentation algorithms that operate over affine transformations and cutout: (1) AutoAugment (Cubuk et al., 2019a), which learns a policy of augmentation methods to apply during training, and (2) Fast AutoAugment (Lim et al., 2019), which is a significantly more efficient version of AutoAugment that does not sub-

Augmentation Method	Accuracy	Policy Search Time (GPU hours)
Crop-and-flip	81.9%	<b>0</b>
Möbius + Cutout	<b>82.9%</b>	<b>0</b>
AutoAugment	<b>82.9%</b>	5,000
Fast AutoAugment	82.7%	3.5

Table 4. Comparison of Möbius with cutout to that of data augmentation policies on CIFAR-100. Our results with Möbius suggest that Möbius performs similarly to these methods already, and can be incorporated into their policy search space for improvement.

stantially sacrifice performance.

Without performing policy search at all, Möbius with cutout achieves the same performance (82.9%) as AutoAugment on CIFAR-100 and outperforms Fast AutoAugment (82.7%). All augmentation methods presented are evaluated on a WRN-28-10. These results are summarized in Table 4.

Our results suggest a couple directions for future work. Most immediately, we plan to add Möbius into the policy search space for an apples-to-apples comparison to these policies. Note that the primary contribution of this work is to introduce Möbius as a successful method for data augmentation on its own and, to a lesser extent, with other sample-level methods such as cutout. Additionally, our results, particularly on low data settings, suggest that Möbius data augmentation could be incorporated into semi-supervised techniques, such as the consistency training objective in UDA (Xie et al., 2019).

## 5. Conclusion

In this paper, we introduce Möbius data augmentation, a method that applies Möbius transformations to images during training to improve model generalization. Empirically, Möbius performs best when applied to a small subset of data in low data settings. Möbius transformations are complementary to other sample-level augmentations that preserve labels, such as cutout or standard affine transformations. In fact, across experiments on CIFAR-10 and CIFAR-100, we find that cutout and Möbius can be combined for superior performance over either alone. In future work, we plan to examine relaxing the constraints on possible Möbius transformations for a more general construction as well as integrating them into the many successful data augmentation policies for fully supervised and semi-supervised learning. Ultimately, this work presents the first foray into successfully employing Möbius transformations—the next level of mathematical abstraction from affine transformations—for data augmentation in neural networks and demonstrating the efficacy of this biologically motivated augmentation on image classification benchmarks.



---

## Appendix A: Defined Möbius Parameters

For Möbius data augmentation, we parameterize eight different Möbius augmentations by specifying two sets of three points  $z_i$  on the original image and  $w_i$  as targets, where  $i = \{1, 2, 3\}$ . Visual examples are shown on different CIFAR-10 classes in Figure 3, which we denote: (1) clockwise twist, (2) clockwise half-twist, (3) spread, (4) spread twist, (5) counter clockwise twist, (6) counter clockwise half-twist, (7) inverse, and (8) inverse spread. Concretely, these parameters are presented below, where  $\Re(p)$  and  $\Im(p)$  denote the respective real and imaginary components of a point  $p$ , and height  $x$  and width  $y$  are dimensions of the original image.

### 1. Clockwise twist:

$$\begin{aligned}\Re(z) &= \{1, 0.5x, 0.6x\}, \\ \Im(z) &= \{0.5y, 0.8y, 0.5y\}, \\ \Re(w) &= \{.5x, 0.5x + 0.3 \sin(0.4\pi)y, 0.5x + 0.1 \cos(0.1\pi)y\}, \\ \Im(w) &= \{y - 1, 0.5y + 0.3 \cos(0.4\pi)y, 0.5y - 0.1 \sin(0.1\pi)x\}\end{aligned}$$

### 2. Clockwise half-twist:

$$\begin{aligned}\Re(z) &= \{1, 0.5x, 0.6x\}, \\ \Im(z) &= \{0.5y, 0.8y, 0.5y\}, \\ \Re(w) &= \{.5x, 0.5x + 0.4y, 0.5x\}, \\ \Im(w) &= \{y - 1, 0.5y, 0.5y - 0.1x\}\end{aligned}$$

### 3. Spread:

$$\begin{aligned}\Re(z) &= \{.3x, 0.5x, 0.7x\}, \\ \Im(z) &= \{0.5y, 0.7y, 0.5y\}, \\ \Re(w) &= \{0.2x, 0.5x, 0.8x\}, \\ \Im(w) &= \{0.5y, 0.8y, 0.5y\}\end{aligned}$$

### 4. Spread twist:

$$\begin{aligned}\Re(z) &= \{.3x, 0.6x, 0.7x\}, \\ \Im(z) &= \{0.3y, 0.8y, 0.3y\}, \\ \Re(w) &= \{0.2x, 0.6x, 0.8x\}, \\ \Im(w) &= \{0.3y, 0.9y, 0.2y\}\end{aligned}$$

### 5. Counter clockwise twist:

$$\begin{aligned}\Re(z) &= \{1, 0.5x, 0.6x\}, \\ \Im(z) &= \{0.5y, 0.8y, 0.5y\}, \\ \Re(w) &= \{0.5x, 0.5x + 0.4y, 0.5x\}, \\ \Im(w) &= \{y - 1, 0.5y, 0.5y - 0.1x\}\end{aligned}$$

### 6. Counter clockwise half-twist:

$$\begin{aligned}\Re(z) &= \{1, 0.5x, 0.6x\}, \\ \Im(z) &= \{0.5y, 0.8y, 0.5y\}, \\ \Re(w) &= \{0.5x, 0.5x + 0.3 \sin(.4\pi)y, 0.5x + 0.1 \cos(.1\pi)x\}, \\ \Im(w) &= \{y - 1, 0.5y + 0.3 \cos(.4\pi)y, 0.5y - 0.1 \sin(.1\pi)x\}\end{aligned}$$

---

7. Inverse:

$$\Re(z) = \{1, 0.5x, x - 1\},$$

$$\Im(z) = \{0.5y, 0.9y, 0.5y\},$$

$$\Re(w) = \{x - 1, 0.5x, 1\},$$

$$\Im(w) = \{0.5y, 0.1y, 0.5y\}$$

8. Inverse spread:

$$\Re(z) = \{0.1x, 0.5x, 0.9x\},$$

$$\Im(z) = \{0.5y, 0.8y, 0.5y\},$$

$$\Re(w) = \{x - 1, 0.5x, 1\},$$

$$\Im(w) = \{0.5y, 0.1y, 0.5y\}$$

---

## References

- Ahlfors, L. *Möbius Transformations in Several Dimensions*. University of Minnesota, 1989.
- Berthelot, D., Carlini, N., Goodfellow, I., Papernot, N., Oliver, A., and Raffel, C. Mixmatch: A holistic approach to semi-supervised learning. *arXiv preprint arXiv:1905.02249*, 2019.
- Criminisi, A., Reid, I., and Zisserman, A. Single view metrology. *International Journal of Computer Vision*, 40(2):123–148, 2000.
- Cubuk, E. D., Zoph, B., Mane, D., Vasudevan, V., and Le, Q. V. Autoaugment: Learning augmentation strategies from data. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 113–123, 2019a.
- Cubuk, E. D., Zoph, B., Shlens, J., and Le, Q. V. Randaugment: Practical data augmentation with no separate search. *arXiv preprint arXiv:1909.13719*, 2019b.
- DeVries, T. and Taylor, G. W. Improved regularization of convolutional neural networks with cutout. *arXiv preprint arXiv:1708.04552*, 2017.
- Ganea, O., Bécigneul, G., and Hofmann, T. Hyperbolic neural networks. In *Advances in neural information processing systems*, pp. 5345–5355, 2018.
- Guo, J. and Gould, S. Deep cnn ensemble with data augmentation for object detection. *arXiv preprint arXiv:1506.07224*, 2015.
- Hernández-García, A. and König, P. Data augmentation instead of explicit regularization. *arXiv preprint arXiv:1806.03852*, 2018.
- Ho, D., Liang, E., Stoica, I., Abbeel, P., and Chen, X. Population based augmentation: Efficient learning of augmentation policy schedules. *arXiv preprint arXiv:1905.05393*, 2019.
- Inoue, H. Data augmentation by pairing samples for images classification. *arXiv preprint arXiv:1801.02929*, 2018.
- Islam, M. A., Anderson, D. T., Pinar, A., Havens, T. C., Scott, G., and Keller, J. M. Enabling explainable fusion in deep learning with fuzzy integral neural networks. *IEEE Transactions on Fuzzy Systems*, 2019.
- Krizhevsky, A., Hinton, G., et al. Learning multiple layers of features from tiny images. 2009.
- Lie, S. Theorie der transformationsgruppen abschn. 3. *Theorie der Transformationsgruppen*, 1893.
- Lim, S., Kim, I., Kim, T., Kim, C., and Kim, S. Fast autoaugment. In *Advances in Neural Information Processing Systems*, pp. 6662–6672, 2019.
- Liouville, J. Extension au cas des trois dimensions de la question du tracé géographique. *Note VI*, pp. 609–617, 1850.
- Lundh, T., Udagawa, J., Hänel, S.-E., and Otani, H. Cross- and Triple-Ratios of Human Body Parts During Development. *Anatomical Record*, 294(8), 2011. ISSN 19328486. doi: 10.1002/ar.21426.
- Mandic, D. P. The use of mobius transformations in neural networks and signal processing. In *Neural Networks for Signal Processing X. Proceedings of the 2000 IEEE Signal Processing Society Workshop (Cat. No. 00TH8501)*, volume 1, pp. 185–194. IEEE, 2000.
- Needham, T. *Visual complex analysis*. Oxford University Press, 1998.
- Özdemir, N., İskender, B. B., and Özgür, N. Y. Complex valued neural network with möbius activation function. *Communications in Nonlinear Science and Numerical Simulation*, 16(12):4698–4703, 2011.
- Perez, L. and Wang, J. The effectiveness of data augmentation in image classification using deep learning. *arXiv preprint arXiv:1712.04621*, 2017.
- Petukhov, S. V. Non-euclidean geometries and algorithms of living bodies. *Computers & Mathematics with Applications*, 17(4-6):505–534, 1989.
- Ratner, A. J., Ehrenberg, H., Hussain, Z., Dunnmon, J., and Ré, C. Learning to compose domain-specific transformations for data augmentation. In *Advances in neural information processing systems*, pp. 3236–3246, 2017.
- Sohn, K., Berthelot, D., Li, C.-L., Zhang, Z., Carlini, N., Cubuk, E. D., Kurakin, A., Zhang, H., and Raffel, C. Fixmatch: Simplifying semi-supervised learning with consistency and confidence. *arXiv preprint arXiv:2001.07685*, 2020.
- Summers, C. and Dinneen, M. J. Improved mixed-example data augmentation. In *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pp. 1262–1270. IEEE, 2019.
- Takahashi, R., Matsubara, T., and Uehara, K. Data augmentation using random image cropping and patching for deep cnns. *IEEE Transactions on Circuits and Systems for Video Technology*, 2019.
- Tang, Z., Peng, X., Li, T., Zhu, Y., and Metaxas, D. N. Adatransform: Adaptive data transformation. In *Proceedings of the IEEE International Conference on Computer Vision*, pp. 2998–3006, 2019.

---

Thompson, D. W. et al. On growth and form. *On growth and form.*, 1942.

Tokozume, Y., Ushiku, Y., and Harada, T. Between-class learning for image classification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 5486–5494, 2018.

Wang, P., Chen, P., Yuan, Y., Liu, D., Huang, Z., Hou, X., and Cottrell, G. Understanding convolution for semantic segmentation. In *2018 IEEE winter conference on applications of computer vision (WACV)*, pp. 1451–1460. IEEE, 2018.

Xie, Q., Dai, Z., Hovy, E., Luong, M.-T., and Le, Q. V. Unsupervised data augmentation. *arXiv preprint arXiv:1904.12848*, 2019.

Zagoruyko, S. and Komodakis, N. Wide residual networks. *arXiv preprint arXiv:1605.07146*, 2016.

Zammit-Mangion, A., Ng, T. L. J., Vu, Q., and Filippone, M. Deep compositional spatial models. *arXiv preprint arXiv:1906.02840*, 2019.

Zhang, H., Cisse, M., Dauphin, Y. N., and Lopez-Paz, D. mixup: Beyond empirical risk minimization. *arXiv preprint arXiv:1710.09412*, 2017.

Zhang, X., Wang, Z., Liu, D., and Ling, Q. Dada: Deep adversarial data augmentation for extremely low data regime classification. In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 2807–2811. IEEE, 2019.

Zhong, Z., Zheng, L., Kang, G., Li, S., and Yang, Y. Random erasing data augmentation. *arXiv preprint arXiv:1708.04896*, 2017.