Visual Chirality

Zhiqiu Lin[1] Jin Sun[1,2] Abe Davis[1,2] Noah Snavely[1,2]
Cornell University[1] Cornell Tech[2]

Introduction

- An object is said to be chiral if it cannot be rotated and translated into alignment with its own reflection, and achiral otherwise.
- We can think of chiral objects as being fundamentally changed by reflection—these are the things that “go the other way” when viewed through a mirror—and we can think of achiral objects as simply being moved by reflection.
- In this paper, the authors define the notion of visual chirality, and analyze visual chirality in real world imagery, both through new theoretical tools, and through empirical analysis.

(a) (b) (c)

Defining visual chirality

- Data augmentation can be seen as a way to improve sampling efficiency for approximating a distribution D(x) (where x represents data from some domain, e.g., images) by assuming that D is invariant to some transformation T

$$\mathbf{D}(\mathbf{x}) = \mathbf{D}(\mathbf{T}(\mathbf{x})) \tag{1}$$

Is a measure of the approximation error associated with assuming visual distributions are symmetric under reflection

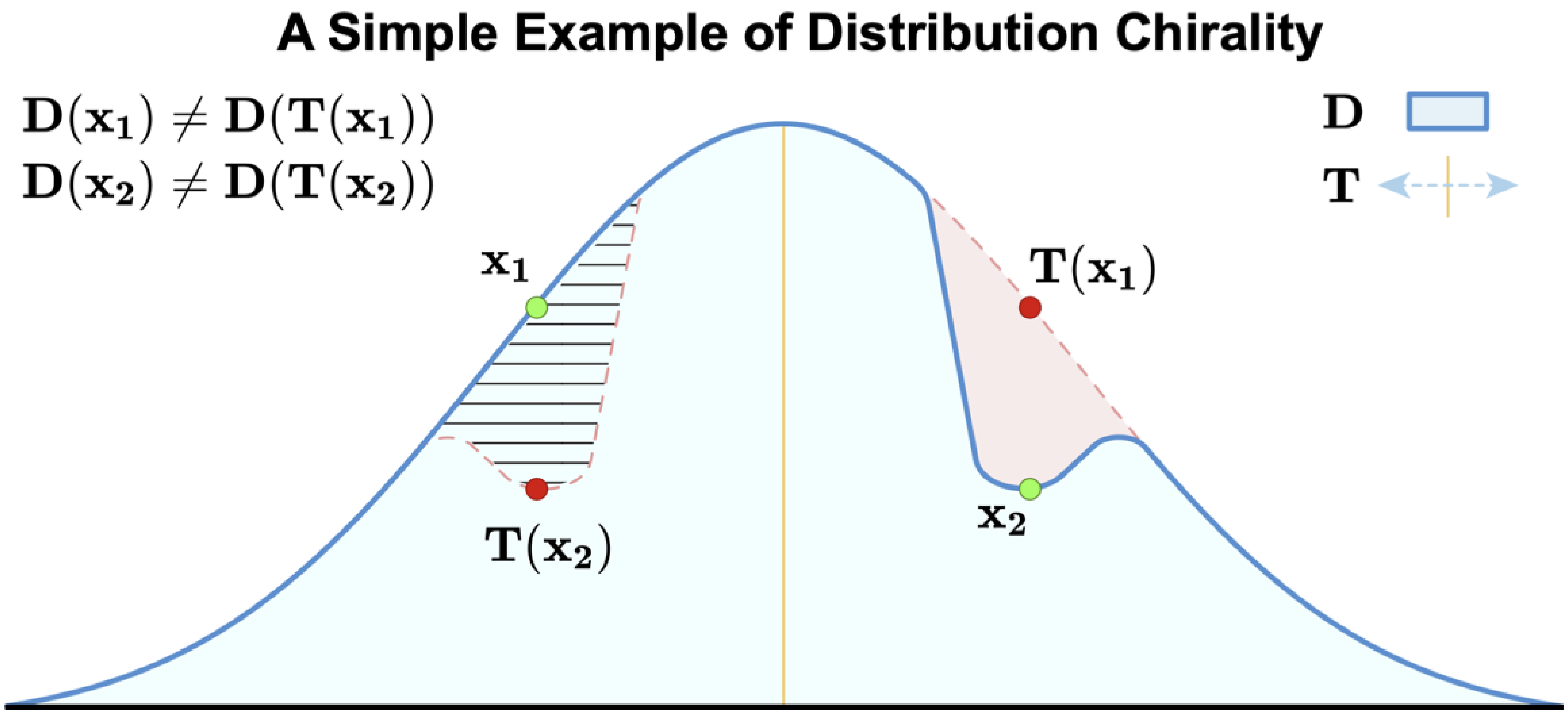

Figure 3. The curve above represents a distribution over images (shown as a 1D distribution for simplicity). Using a transformation $\mathbf{T}$ to augment a sample-based approximation of the distribution $\mathbf{D}$ assumes symmetry with respect to $\mathbf{T}$. We define visual chirality in terms of approximation error induced by this assumed symmetry when $\mathbf{T}$ is image reflection.

Measuring visual chirality

- The authors denote a set of training images from some distribution as
  - $C_{positive} = \{I_1, I_2, \cdots, I_n\}$
- They perform a horizontal flip on each image I_i to produce its reflected version I_i'. Let us denote the mirrored set as:
  - $C_{\text{negative}} = \{I_1', I_2', \cdots, I_n'\}$
- they then assign a binary label y_i to each image I_i in $C_{\text{positive}} \cup C_{\text{negative}}$:

$$y_i = \begin{cases} 0 & \text{if } I_i \in C_{\text{negative}}, \text{ i.e., flipped} \\ 1 & \text{if } I_i \in C_{\text{positive}}, \text{ i.e., non-flipped} \end{cases} \qquad (2)$$

The chirality of image processing

Term	Definition	Meaning in Learning Applications
A distribution $\mathbf{D}$	$\mathbf{D} : \mathbb{R}^n \mapsto \mathbb{R}$	The underlying distribution our training data is drawn from for some task.
A symmetry transformation $\mathbf{T}$	$\mathbf{T} : \mathbb{R}^n \mapsto \mathbb{R}^n$, is associative and invertible	E.g., horizontal reflection, or any other associative and invertible transformation to be used for data augmentation.
A processing transformation $\mathbf{J}$	$\mathbf{J} : \mathbb{R}^n \mapsto \mathbb{R}^n$, does *not* have to be invertible	Some combination of image processing operations, e.g., demosaicing and/or JPEG compression.
A transformed distribution $\mathbf{D_J}$	$\mathbf{D_J}(\mathbf{y}) = \sum_{\mathbf{x} \in \mathbf{J}^{-1}(\mathbf{y})} \mathbf{D}(\mathbf{x})$	The distribution of training data after every element has been transformed by $\mathbf{J}$.

Table 1. Terms and definitions used in derivations.

The chirality of image processing

- **Transformation commutativity**
  - The first is simply to say that if some symmetry exists in the distribution **D** then the same symmetry should exist in **DJ**, the transformation of that distribution by J
  - if elements x_a and x_b are related by $x_b = \mathbf{T}x_a$, then this relationship should be preserved by **J**, meaning $\mathbf{J}x_b = \mathbf{TJ}x_a$

The chirality of image processing

- **Predicting chirality with commutativity**
  - Demosaicing and JPEG compression are both individually chiral and chiral when combined.
  - When random cropping is added to demosaicing or JPEG compression individually, they become achiral.
  - When demosaicing, JPEG compression, and random cropping are all combined, the result is chiral.

High-level visual chirality

High-level visual chirality

Figure 5. **Chiral clusters discovered in the Instagram dataset.** Each row shows selected images from a single discovered cluster. Each

Visual chirality in faces

Figure 6. **Chiral clusters found in FFHQ.** It shows 3 chiral clusters of FFHQ dataset. The leftmost image of each row is the average face + CAM heatmap for all non-flipped images inside the each cluster. We also show some random non-flipped examples for each cluster.

Datasets

- **StreetStyle dataset:**
  - consists of millions of images of people gathered from Instagram. For our work, we select a random subset of 700K images from StreetStyle, and refer to this as the *Instagram* dataset
- **Flickr100M**
  - Randomly selected subset
- **FFHQ dataset**
  - FFHQ is a recent dataset of 70K high-quality faces introduced in the context of training generative methods.

Results

Training set	Preprocessing	Test Accuracy	
		Instagram	F100M
Instagram	Resizing	0.92	0.57
Instagram	RandCrop	0.80	0.59
Instagram (no-text)	RandCrop	0.74	0.55

Table 1. **Chirality classification performance of models trained on *Instagram*.** Hyper-parameters were selected by cross validation. The first column indicates the training dataset, and the second column the processing that takes place on input images. The last columns report on a held-out test set, and on an unseen dataset (Flickr100M, or F100M for short). Note that the same preprocessing scheme (resize vs. random crop) is applied to both the training and test sets, and the model trained on *Instagram* without text is also tested on *Instagram* without text.

Conclusion

- The author proposed an algorithm to discover visual chirality in image distributions using a self-supervised learning approach.
- They achieve this by predicting whether a photo is flipped or not, and by analyzing properties of transformations that yield chirality.
- Their results implies that visual chirality indeed exists in many vision datasets.