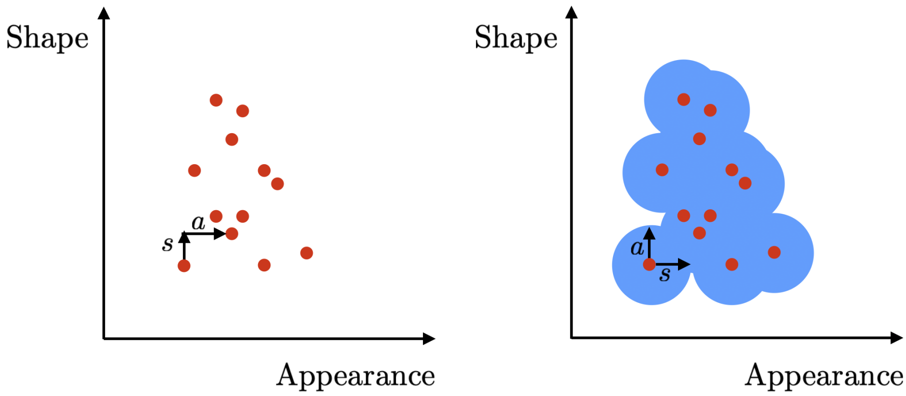
1 Object Shape Learning

1.1 Transformations



(a) (b)

Figure 1.1: Effect of transformations on data distribution: (a) Data points (red) can be connected via a shape s and an appearance a transformation. (b) Applying transformations effectively blurs the data distribution.

In this section we discuss the effect of the transformations on learning a consistent and comprehensive representation. Since strong image transformations can make the learning curve for the algorithm too steep, we exponetially schedule the increase in magnitude, finally resulting in image changes as shown in Fig. **??**. In effect, the transformations teach the algorithm what changes in shape and appearance are. Assuming that samples from the data distribution are - showing the same object class - related via a change in shape and appearance, the transformations blur the distribution. This data augmentation is sketched in Fig. 1.1.

1.1.1 Spatial Transformations

We perform thin-plate spline (TPS) warps to mimick spatial transformations. These changes incorporate rotation, scaling and translation as a special case. While irreplacable for calculating the direct equivariance loss, they can result in artificial shape changes. After all, most objects - such as human beings or animals, do not warp, but articulate their parts/limbs. Natural shape changes are needed to learn a model of the objects articulation. These changes are presented in video data. Hence, for videos we enforce the reconstruction to function across different frames. This results in a much stabler performance and greater part consistency especially for highly articulated parts such as arms.

1.1.2 Appearance Transformations

Figure 1.2: Examples for shape and appearance transformation on CUB-200-2011.

We mimick appearance changes with image transformations in color, contrast, hue and brightness. Exemplars for the combined effect of spatial and appearance transformations are shown in Fig. **??**. Ideally one would want

on Cats -> black cats different set of KP than rest -> connect these samples via transformation to reach intra-class consistency

1.1.3 Parity

birds parity salsa parity Since the human body appearances in the frontal and back views are similar, we do not expect our discovered landmarks to distinguish the left and right sides of the human body, which means that a landmark at the left leg in the frontal view can locate the right leg in the back view.

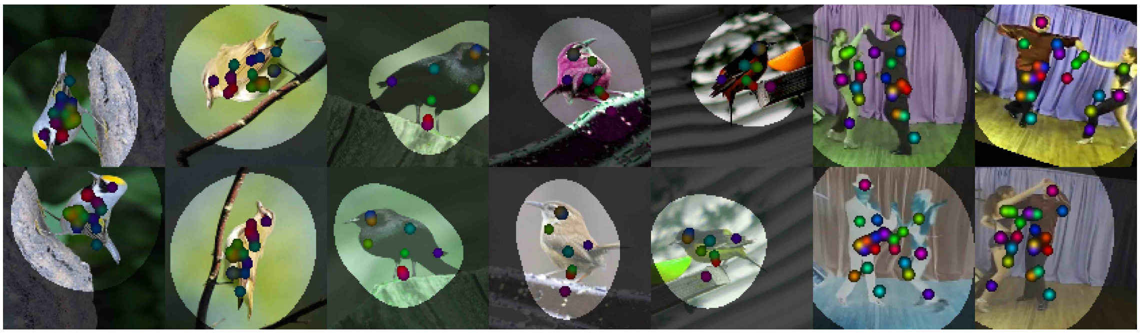


Figure 1.3: Parity changes: the images of the upper and lower row relate via the usual transformations and an additional parity flip. For the bird (1-5th column) images induced artificially, for the dancing humans (6-7th column) via sampling different frames from a video.

2 Bibliography