SpecMix: audio data augmentation method

Pavel Lukianov

Moscow, 2021

Image Augmentation

Augmentation - method used to increase the amount of data by adding modified copies of already existing data or created synthetic data from existing data.

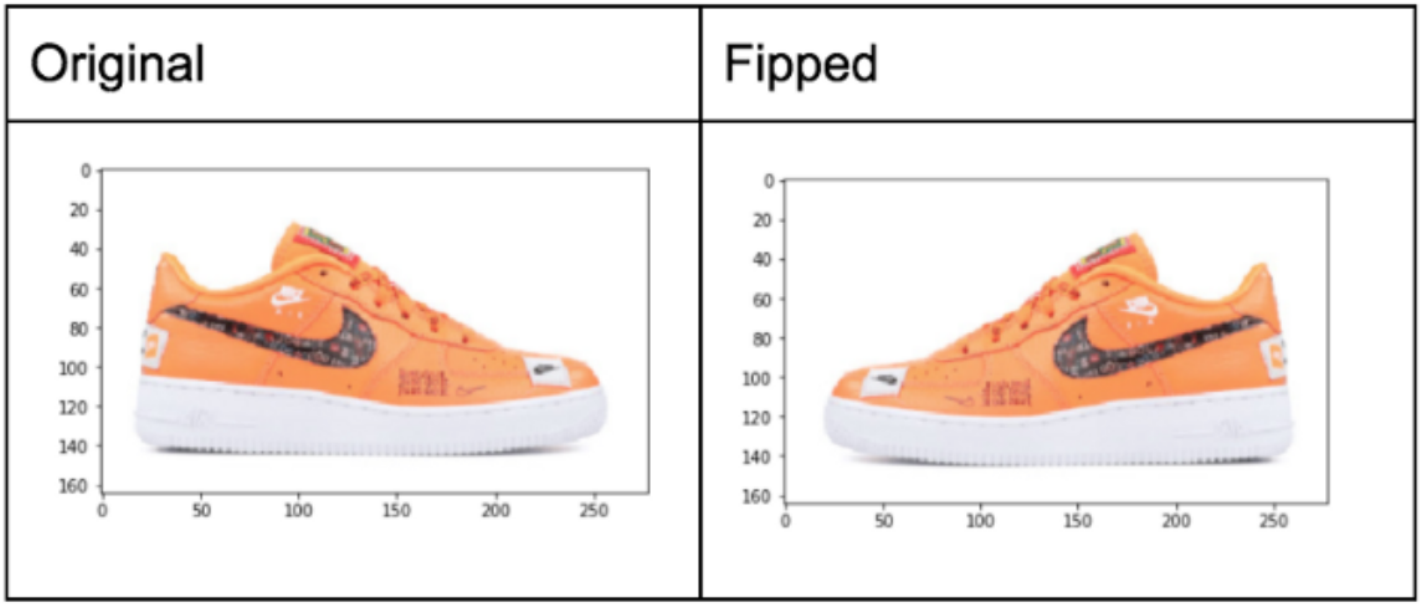

Figure: Flipping

SpecAugment

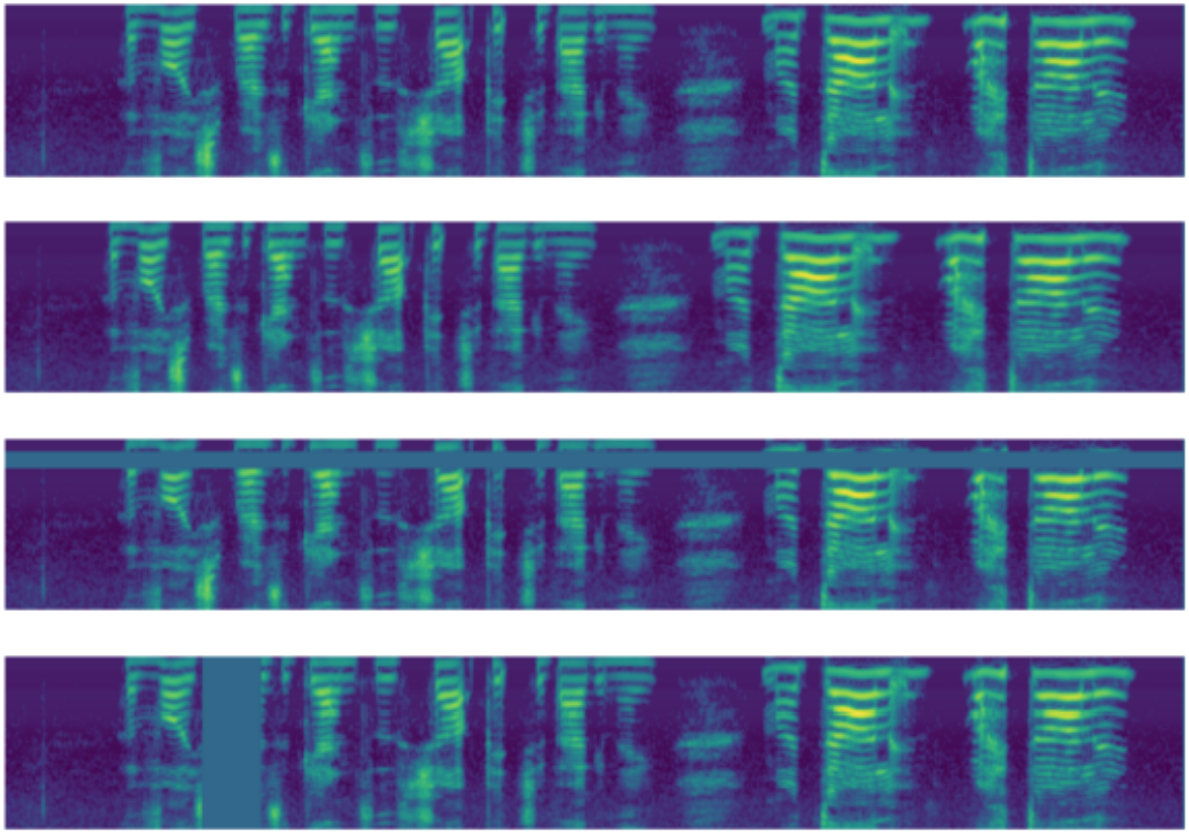

Figure: Original, Time warping, Frequency masking, Time masking

Mixup

There are two training samples (x_i, y_i), (x_j, y_j)

$x_{new} = \lambda x_i + (1 - \lambda) x_j$
$y_{new} = \lambda y_i + (1 - \lambda) y_j$, where $\lambda \sim Beta(\alpha, \alpha)$

Linear interpolations of feature vectors should lead to linear interpolations of the associated targets.

Cutmix

There are two training samples (x_i, y_i), (x_j, y_j)

$x_{new} = M \odot x_i + (1 - M) \odot x_j$
$y_{new} = \lambda y_i + (1 - \lambda) y_j$, where $\lambda \sim Beta(\alpha, \alpha)$

Bounding box coordinates $B = (r_x, r_y, r_w, r_h)$:

$r_x \sim Unif(0, W)$, $r_w = W\sqrt{1 - \lambda}$
$r_y \sim Unif(0, H)$, $r_h = H\sqrt{1 - \lambda}$

$\frac{r_w r_h}{HW} = 1 - \lambda$

Mixup and Cutmix

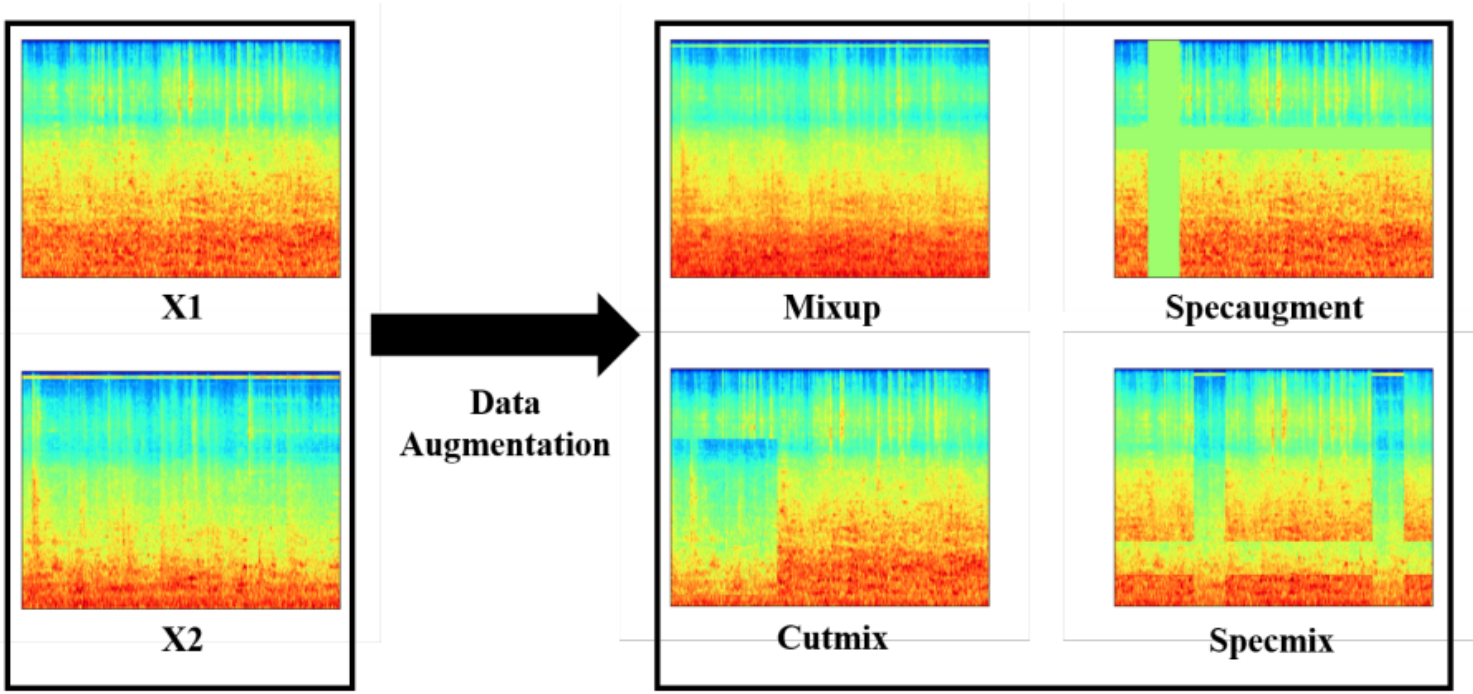

Figure: Mixup, Cutmix, Specaugment and Specmix

SpecMix

There are two training samples (x_i, y_i), (x_j, y_j)

$$x_{new} = M \odot x_i + (1 - M) \odot x_j$$
$$y_{new} = \lambda y_i + (1 - \lambda) y_j$$

λ - share of x_i in x_{new}

Masking strategy:

- Time masking
- Frequency masking

SpecMix

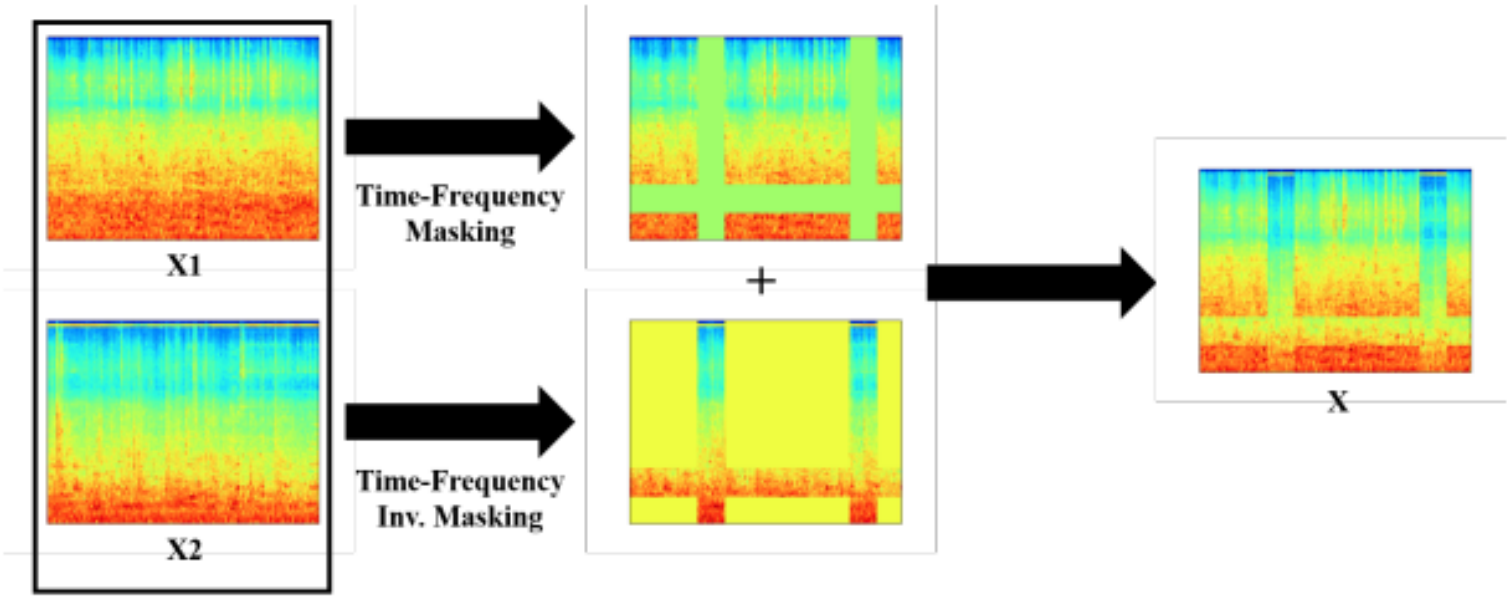

Figure: SpecMix

Experiments

- cross-entropy loss
- batch_size = 32

Model : ResNet-101	Accuracy(%)
No augmentation	59.60
Mixup [15]	58.15
Cutmix [18]	59.54
Specaugment [11]	57.68
Specmix $\gamma = 0.3$	**62.13**

Figure: TAU Urban Acoustic Scenes 2020 Mobile benchmark

Experiments

- cross-entropy loss
- batch_size = 32

Model : [5]	Accuracy(%)
No augmentation	68.90
Mixup [15]	71.29
Cutmix [18]	70.92
Specaugment [11]	69.07
Specmix $\gamma = \mathcal{U}[0, 1]$	**71.60**

Figure: TAU Urban Acoustic Scenes 2020 Mobile benchmark

Experiments

- cross-entropy loss
- batch_size = 32

Model : ResNet-101	Accuracy(%)
No augmentation	96.07
Mixup [15]	95.87
Cutmix [18]	95.66
Specaugment [11]	96.90
Specmix $\gamma = \mathcal{U}[0, 1]$	**97.11**

Figure: SECL UMONS benchmark