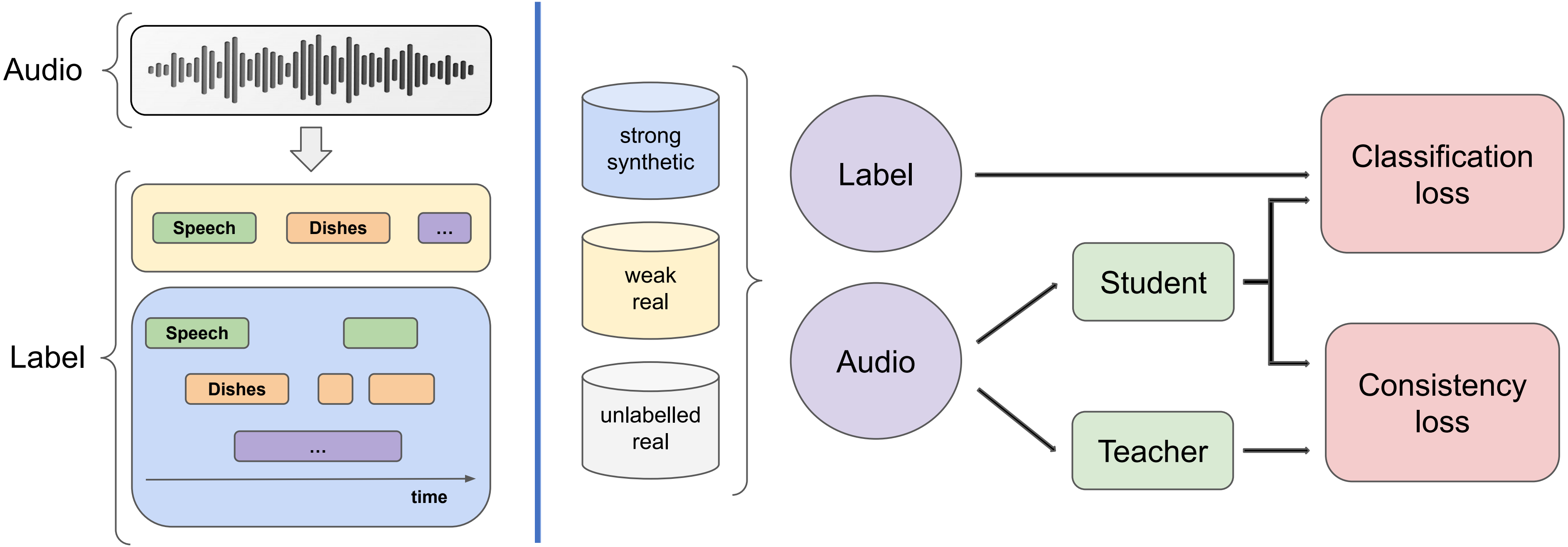
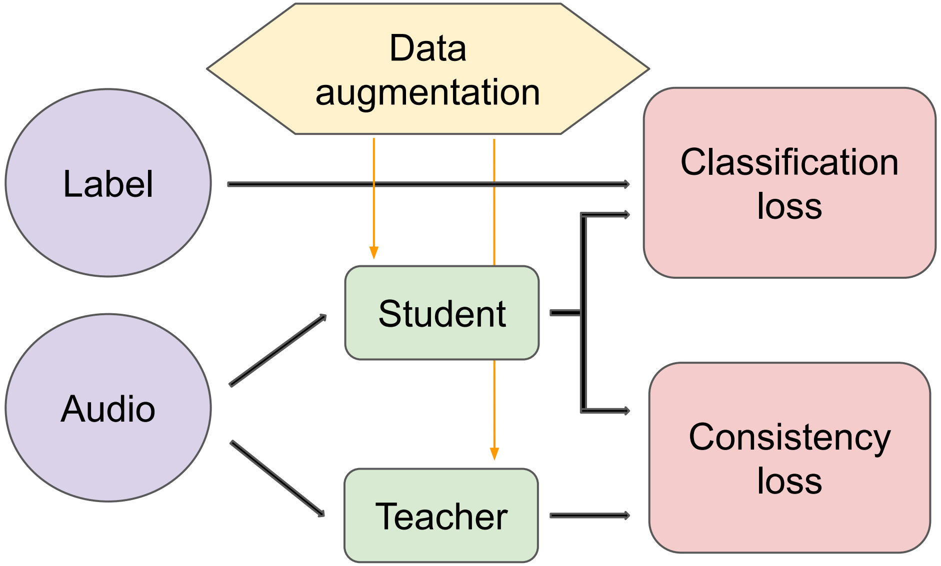
Adversarial and latent data augmentation for sound event detection

David Perera, Slim Essid, Gaël Richard

Learning invariants for DCASE Task 4



Data augmentation strategies



Consistency loss

$$\mathcal{L}_{consistency} = ||f(x) - f(x+d)||$$

Augmentation strategies

$$d_{random} \sim N(0, I)$$

$$d_{adversarial} = \nabla_d ||f(x) - f(x+d)||$$

$$d_{VAT} = argmax_{||d|| \leq \epsilon} ||f(x) - f(x+d)||$$

$$d_{Mixup} = Mixup(x, x')$$

Contributions

- We compare the impact of several data augmentation strategies on DCASE task 4.
- We propose a training objective that is more **flexible**, requires less **gpu memory**, is easier to **interpret** and leads to better **performance** than task 4 baseline.
- The performance gain is **class invariant** and resilient to a **diminution of the volume of training data**