Introduction

- Tackle privacy risks encountered in Acoustic Sensor Network applications
- Illustrate concept with a smart office and challenging competing goals scenario
- Balance competing goals: utility (gender discrimination) & privacy (speaker identification)

Defender vs. Attacker

- Previous work [1]:
  - traditional feature representation carries significant speaker-dependent data
  - adversarial feature extraction successfully used but depends on attacker configuration
- More general approach: privacy-aware variational information feature extraction:
  - inspired by variational information autoencoders [2] which use information minimization
  - the encoding variable is a compact stochastic feature representation
  - the proposed system is described in Fig. 1

Train defender

- \[ \min_{\theta_f, \theta_g, \beta} \mathbb{E}_{p(x|z)}[\log p(x|z)] + \beta I(X; Z) \] (1)
- \( \theta_f, \theta_g, \beta \) are weights and biases; \( \Gamma \) and \( f \) are true and predicted gender labels
- \( I(X; Z) \) is the mutual information between input set \( X \) and encoding set \( Z \)
- \( \beta \) is a budget scaling factor for controlling information minimization
- \( I(X; Z) \) is computationally challenging, find analytical upper bound \( I_{\text{max}}(X; Z) \geq I(X; Z) \):
  - \[ I(X; Z) = \int p(z|x)\log p(z|x) \, dz - \int p(z)\log p(z) \, dz \] (2)
- construct encoding variable \( z = \sigma(x) = \epsilon + \mu(x) \), where \( \epsilon \sim N(0, 1) \)
- now \( p(z|x) \) follows a Gaussian distribution \( N(\mu(x), \sigma(x)) \)
- backpropagation can be efficiently performed by updating \( \theta_f \) and \( \theta_g \) [3]

Experimental Results

- Speaker identification risks can be drastically reduced without significantly deteriorating gender discrimination accuracy
- Each input \( X \) gets mapped to a distribution rather than a unique \( Z \) which in turn, controlled by \( \beta \), ignores as many details of \( X \) as possible
- Proposed concept can be further expanded to other utility vs. privacy applications

Conclusions and Outlook

References