**Introduction**

- Domain mismatch poses a great challenge to speaker verification (SV). Domain adaptation is often adopted to overcome this problem.
- Domain adversarial neural network (DANN) is a state-of-the-art domain adaptation method for SV. It uses a speaker classifier and a domain discriminator to learn speaker discriminative and domain-invariant features.
- Limitation of DANN: there is no guarantee that the learned features follow a Gaussian distribution, which is an essential requirement for the Gaussian PLDA backend.

**Variational Domain-Adversarial Neural Network (VDANN)**

- **Methodology:** incorporate a variational auto-encoder (VAE) into the DANN to impose constraint on the distribution of the embedded features.
- **Objective:** produce features that are not only speaker discriminative and domain-invariant but also Gaussian distributed.
- **Architecture:** VDANN comprises a speaker predictor $c$, a domain classifier $D$, an encoder $E$ and a decoder $G$. Their parameters are denoted as $\theta_c$, $\theta_D$, $\phi_E$ and $\theta_G$, respectively.
- **Optimization:**
  - Keeping $\theta_c$, $\phi_E$ and $\theta_G$ fixed, minimize the domain classification loss with respect to $\theta_D$;
  - Keeping $\theta_D$ fixed, maximize the domain classification loss while simultaneously minimizing the speaker classification loss and the VAE loss with respect to $\theta_c$, $\phi_E$ and $\theta_G$.

**Experimental Setup**

- X-vectors were extracted using the pre-trained DNN available from the Kaldi repository.
- We trained the VDANN/DANN on SRE04–10, Voxceleb1, Switchboard 2 Phases I–III and SITW datasets. The DANN has the same structure as the VDANN, but without the VAE decoder and the sampling procedure.
- The baseline PLDA model was trained on the SRE04–10 and their augmented x-vectors for SRE16; while for SRE18 the Mix6 and its augmented x-vectors were also added to the training sets. For VDANN/DANN evaluation, the PLDA model was trained on the transformed x-vectors.
- Pre-processing includes centering and LDA projection (to a 150 dimensional space).

**References**