Multimodal Factorization Model

- **Bayesian Network**
- **Generative Network**
- **Inference Network**

### Notations
- $X_{ij}$: multimodal data from $M$ modalities, $Y$: labels
- $F$: generated multimodal data, $Y$: generated labels
- $Z$: modality-specific latent variables, $F$: factors

### Summary
- Joint generative-discriminative objective for multimodal data
- Factorize representation into independent sets of factors
- Multimodal Discriminative factors
- Modality-Specific Generative factors

### Neural Architecture
- Encoder $Q$ forms multimodal fusion.
- Inference Network
- Generative Network

### Contributions
- SOTA performance on six multimodal datasets
- Flexible generative capabilities by independent factors.
- Ability to reconstruct missing modalities.
- Interpreting multimodal interactions.

### Generation, Inference, and Learning

#### Generation
- Joint distribution Wasserstein distance
- Approximation for intractable exact inference

#### Inference
- Joint-Distribution Wasserstein Distance

#### Relaxed Objective

#### Surrogate Inference for Missing Modalities

#### Controllable Generation

### Details Dataset
- Handwritten (MNIST) + Street-view House Numbers (SVHN)

### Results
- **Ablation Study**
  - On CMU-MOSI
  - **Missing Modalities**
    - On CMU-MOSI
    - **Neural Architecture**
      - Joint generative-discriminative objective for multimodal data.
- **SOTA performance on six multimodal datasets.**
- **Encoder**
- **Inference Network**
- **Generative Network**

### References