Improving Multimodal Fusion Via Mutual Dependency Maximisation

Oral Presentation at EMNLP 2021

Pierre Colombo, Emile Chapuis
Matthieu Labeau, Chloé Clavel

What is our definition of multimodality?

Slide inspired from: https://www.cs.cmu.edu/~morency/MMML-Tutorial-ACL2017.pdf

What is our definition of multimodality?

Verbal

« **What you say ?** »

- Lexicon:
  - Words
- Syntax:
  - POS
- Pragmatics:
  - DA
  - Emotion

Slide inspired from: https://www.cs.cmu.edu/~morency/MMML-Tutorial-ACL2017.pdf

What is our definition of multimodality?

Verbal

« What you say ? »

- Lexicon:
  - Words
- Syntax:
  - POS
- Pragmatics:
  - DA
  - Emotion

« How you say it ? »

Slide inspired from: https://www.cs.cmu.edu/~morency/MMML-Tutorial-ACL2017.pdf

What is our definition of multimodality?

Verbal

« What you say ? »

- Lexicon:
  - Words
- Syntax:
  - POS
- Pragmatics:
  - DA
  - Emotion

« How you say it ? »

Vocal

- Prosody
  - Intonation
  - Voice quality
- Vocal expressions:
  - Laughter
  - Moans

Slide inspired from: https://www.cs.cmu.edu/~morency/MMML-Tutorial-ACL2017.pdf

What is our definition of multimodality?

Verbal

« What you say ? »

- Lexicon:
  - Words
- Syntax:
  - POS
- Pragmatics:
  - DA
  - Emotion

« How you say it ? »

Vocal

- Prosody
  - Intonation
  - Voice quality
- Vocal expressions:
  - Laughter
  - Moans

Visual

- Gestures:
  - Head & Eye & Arm
- Body language
  - Body posture
  - Proxemics
- Eye contact
  - Head & Eye gaze
- Facial expressions
  - FACS action units
  - Smile & Frowning

Slide inspired from: https://www.cs.cmu.edu/~morency/MMML-Tutorial-ACL2017.pdf

Core Challenges in Multimodal Learning

Slide inspired from: https://www.cs.cmu.edu/~morency/MMML-Tutorial-ACL2017.pdf

Core Challenges in Multimodal Learning

5 challenges of multimodal learning

Slide inspired from: https://www.cs.cmu.edu/~morency/MMML-Tutorial-ACL2017.pdf

Core Challenges in Multimodal Learning

5 challenges of multimodal learning

Representation

Represent multimodal data (leverage complementarity, redundancy)

Core Challenges in Multimodal Learning

5 challenges of multimodal learning

Representation	**Alignement**
Represent multimodal data (leverage complementarity, redundancy)	**Identify** relations between elements of different modalities

Slide inspired from: https://www.cs.cmu.edu/~morency/MMML-Tutorial-ACL2017.pdf

Core Challenges in Multimodal Learning

5 challenges of multimodal learning

Representation	Alignement	Fusion
Represent multimodal data (leverage complementarity, redundancy)	**Identify** relations between elements of different modalities	**Join** information from modalities

Slide inspired from: https://www.cs.cmu.edu/~morency/MMML-Tutorial-ACL2017.pdf

Core Challenges in Multimodal Learning

5 challenges of multimodal learning

Representation	Alignement	Fusion	Translation
Represent multimodal data (leverage complementarity, redundancy)	**Identify** relations between elements of different modalities	**Join** information from modalities	**Translate** one modality to another

Slide inspired from: https://www.cs.cmu.edu/~morency/MMML-Tutorial-ACL2017.pdf

Core Challenges in Multimodal Learning

5 challenges of multimodal learning

Representation → **Represent** multimodal data (leverage complementarity, redundancy)

Alignement → **Identify** relations between elements of different modalities

Fusion → **Join** information from modalities

Translation → **Translate** one modality to another

Co-learning → **Transfer** knowledge between modalities

Slide inspired from: https://www.cs.cmu.edu/~morency/MMML-Tutorial-ACL2017.pdf

Core Challenges in Multimodal Learning

5 challenges of multimodal learning

Representation	Alignement	Fusion	Translation	Co-learning
Represent multimodal data (leverage complementarity, redundancy)	Identify relations between elements of different modalities	Join information from modalities	Translate one modality to another	Transfer knowledge between modalities

Slide inspired from: https://www.cs.cmu.edu/~morency/MMML-Tutorial-ACL2017.pdf

Concurrent Works

Concurrent Works

Previous Work on $\boxed{\text{multimodal sentiment analysis}}$

Model		Fusion Mechanism
TFN	Zadeh et al. 2017	Tensor fusion
MARN	Zadeh et al. 2018	Multi attention block
LFN	Liu et al. 2018	Low rank tensor fusion
MFN	Zadeh et al 2018	Delta Memory attention network
MISA	Hazarika et al. 2020	Transformer - Multihead
MAGBERT/MAGXLNET	Rahman et al. 2020	Multimodal Adaptation Gate

Concurrent Works

Previous Work on multimodal sentiment analysis

Model		Fusion Mechanism
TFN	Zadeh et al. 2017	Tensor fusion
MARN	Zadeh et al. 2018	Multi attention block
LFN	Liu et al. 2018	Low rank tensor fusion
MFN	Zadeh et al 2018	Delta Memory attention network
MISA	Hazarika et al. 2020	Transformer - Multihead
MAGBERT/MAGXLNET	Rahman et al. 2020	Multimodal Adaptation Gate

Fusion in previous work is mainly neural
Few previous works on loss function !

Concurrent Works

Previous Work on multimodal sentiment analysis

Model		Fusion Mechanism
TFN	Zadeh et al. 2017	Tensor fusion
MARN	Zadeh et al. 2018	Multi attention block
LFN	Liu et al. 2018	Low rank tensor fusion
MFN	Zadeh et al 2018	Delta Memory attention network
MISA	Hazarika et al. 2020	Transformer - Multihead
MAGBERT/MAGXLNET	Rahman et al. 2020	Multimodal Adaptation Gate

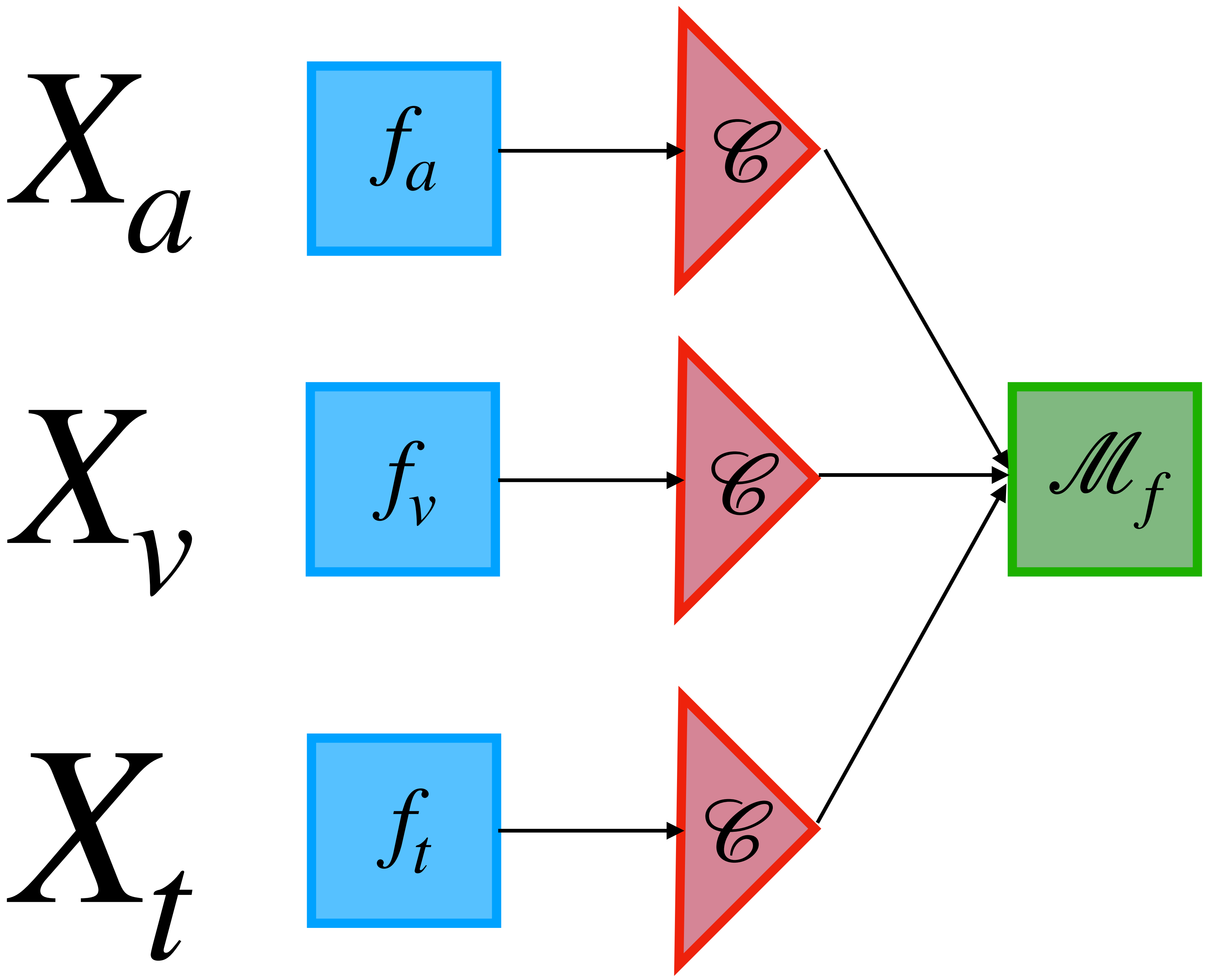
Fusion in previous work is mainly neural
Few previous works on loss function !

Contribution: A new class of loss functions $\mathcal{L}_{MDM} \triangleq MDM\left(p_{X_a X_v X_l}(x_a, x_v, x_l), \prod_{j\in\{a,v,l\}} p_{X_j}(x_j)\right)$

Background in Multimodal sentiment analysis

Background in Multimodal sentiment analysis

Late Fusion

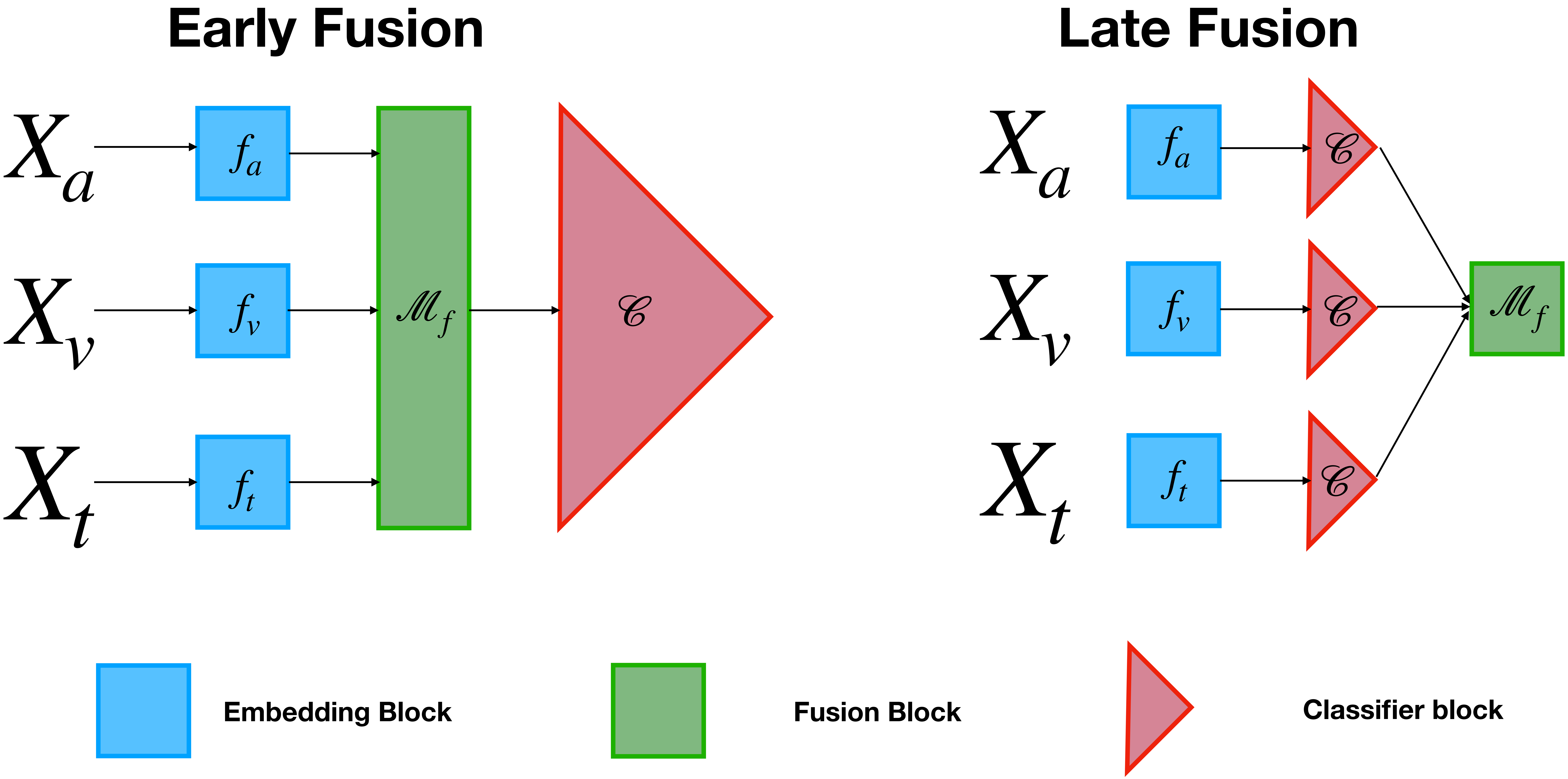


Embedding Block

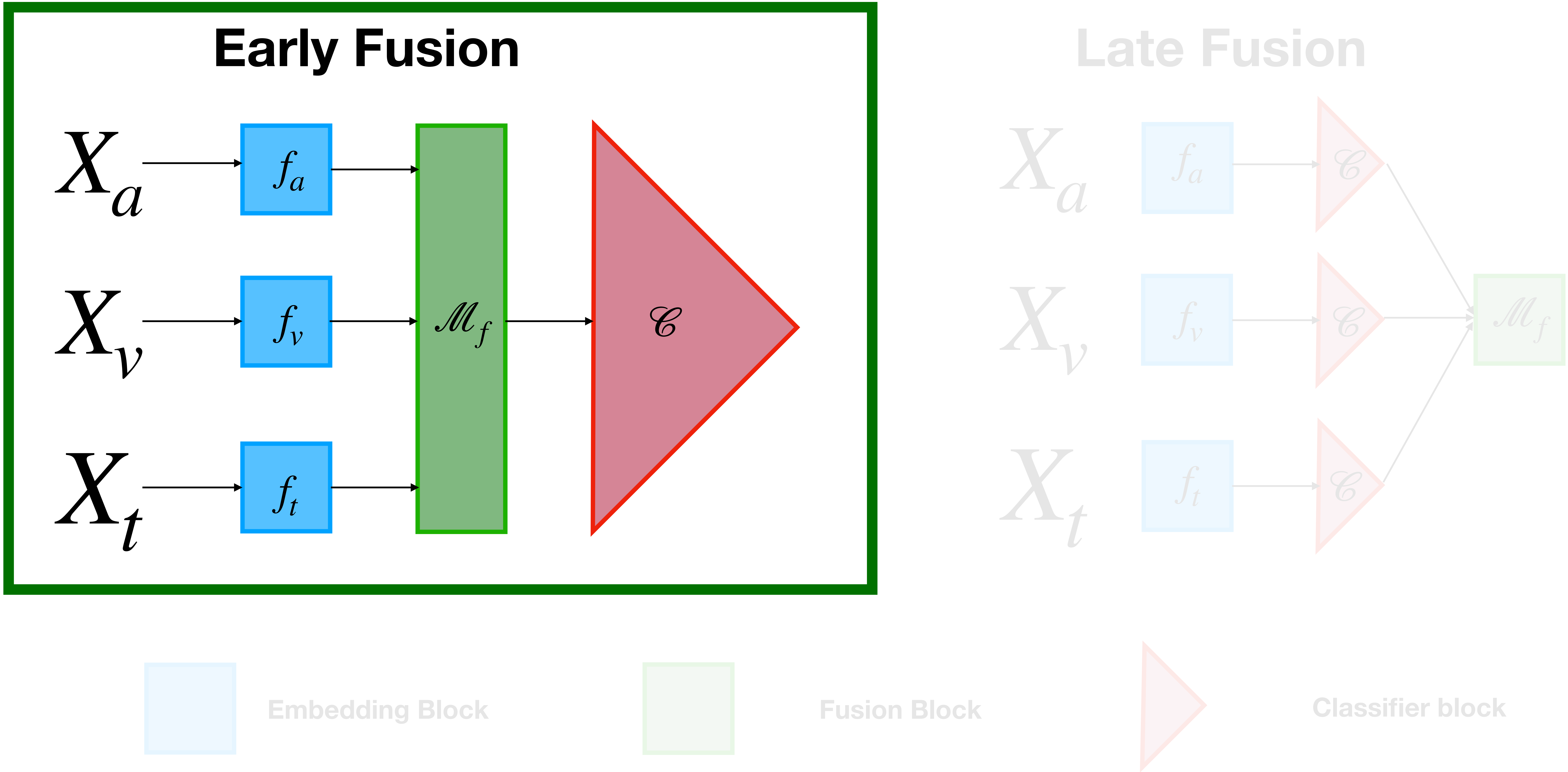
Fusion Block

Classifier block

Background in Multimodal sentiment analysis



Background in Multimodal sentiment analysis



Problem formulation

Problem formulation

Fusion Mechanism $\mathscr{M}_f : \mathscr{X}_a \times \mathscr{X}_v \times \mathscr{X}_l \rightarrow \mathscr{R}^d$

Problem formulation

Fusion Mechanism $\mathscr{M}_f : \mathscr{X}_a \times \mathscr{X}_v \times \mathscr{X}_l \to \mathscr{R}^d$

What we want for $\mathscr{M}_f$:

- **Retain modality-specific interaction**
- **Retain cross-view interaction**
- **Retain task specific information**

$\mathscr{L}_{MDM}$

$\mathscr{L}_{down.}$

Problem formulation

Fusion Mechanism $\mathcal{M}_f : \mathcal{X}_a \times \mathcal{X}_v \times \mathcal{X}_l \rightarrow \mathcal{R}^d$

What we want for $\mathcal{M}_f$:

- **Retain modality-specific interaction** $\mathcal{L}_{MDM}$
- **Retain cross-view interaction**
- **Retain task specific information** $\mathcal{L}_{down.}$

Total $\underbrace{\mathcal{L}_{down.}}_{\text{main task}} - \underbrace{\lambda \cdot \mathcal{L}_{MDM}}_{\text{mutual dependency term}}$

$$\mathcal{L}_{MDM} \triangleq MDM\left(p_{X_a X_v X_l}(x_a, x_v, x_l), \prod_{j\in\{a,v,l\}} p_{X_j}(x_j)\right)$$

Problem formulation

Fusion Mechanism $\mathcal{M}_f : \mathcal{X}_a \times \mathcal{X}_v \times \mathcal{X}_l \to \mathcal{R}^d$

What we want for $\mathcal{M}_f$**:**

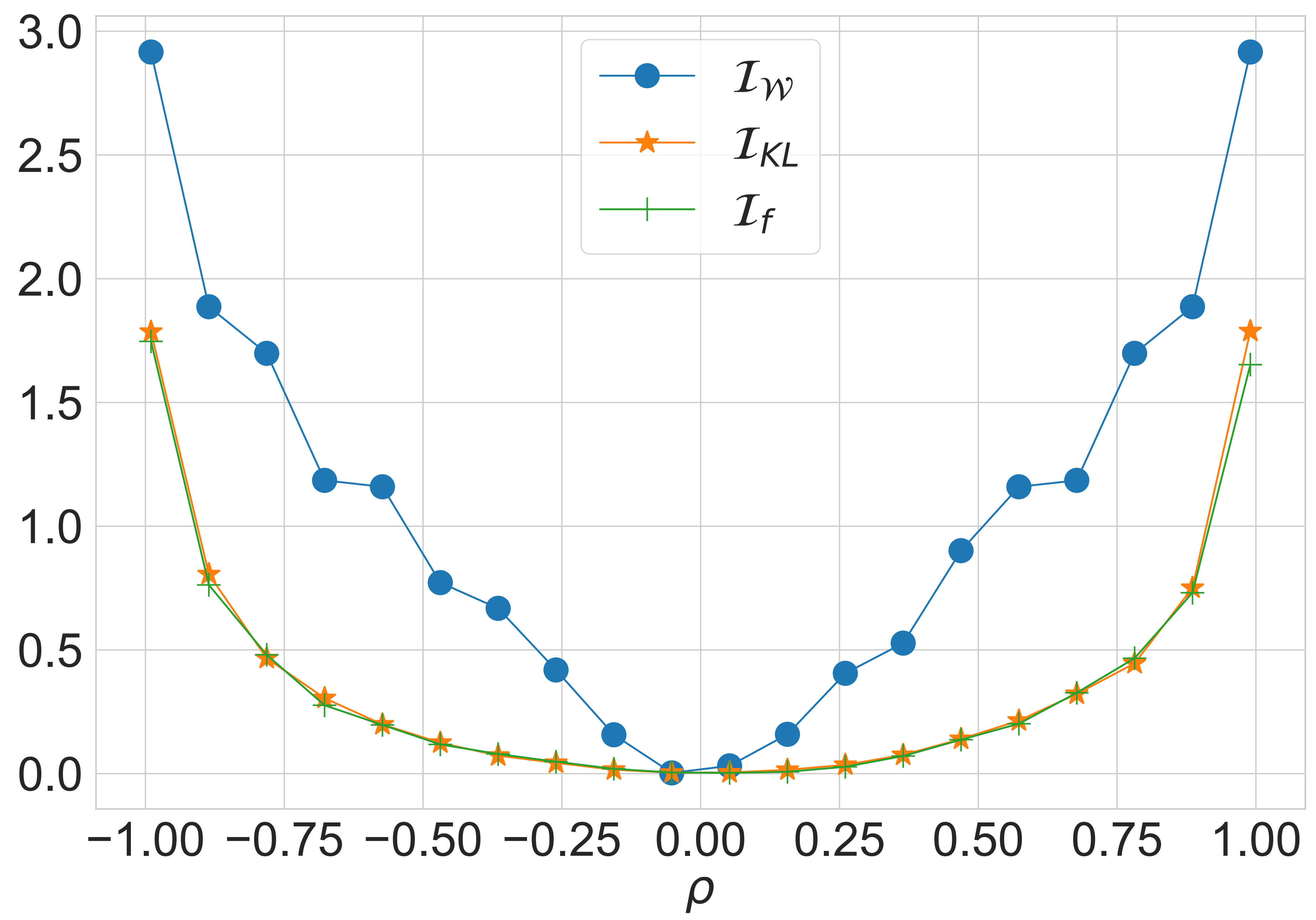
- **Retain modality-specific interaction** $\mathcal{L}_{MDM}$
- **Retain cross-view interaction**
- **Retain task specific information** $\mathcal{L}_{down.}$

Total $\underbrace{\mathcal{L}_{down.}}_{\text{main task}} - \underbrace{\lambda \cdot \mathcal{L}_{MDM}}_{\text{mutual dependency term}}$

$$\mathcal{L}_{MDM} \triangleq MDM\left(p_{X_a X_v X_l}(x_a, x_v, x_l), \prod_{j \in \{a,v,l\}} p_{X_j}(x_j)\right)$$

Different choice for MDM

- **Wasserstein**
- **Kullback Leibler (KL)**
- **F-divergence**



Toy example on correlated Gaussian

Training Losses

Training Losses

Training Loss

$$\mathscr{L} \triangleq \underbrace{\mathscr{L}_{down.}}_{\textbf{main task}} - \underbrace{\lambda \cdot \mathscr{L}_{MDM}}_{\textbf{mutual dependency term}}$$

Training Losses

Training Loss

$$\mathscr{L} \triangleq \underbrace{\mathscr{L}_{down.}}_{\textbf{main task}} - \underbrace{\lambda \cdot \mathscr{L}_{MDM}}_{\textbf{mutual dependency term}}$$

Notation

$$X_a \sim p_{X_a}, X_v \sim p_{X_v}, X_l \sim p_{X_l} \qquad X_a, X_v, X_l \sim p_{X_a, X_v, X_l}$$

$$T(\theta) : \mathscr{X}_a \times \mathscr{X}_v \times \mathscr{X}_l \to \mathbb{R}$$

family of functions

Training Losses

Training Loss $$\mathscr{L} \triangleq \underbrace{\mathscr{L}_{down.}}_{\text{main task}} - \underbrace{\lambda \cdot \mathscr{L}_{MDM}}_{\text{mutual dependency term}}$$

Notation $$X_a \sim p_{X_a}, X_v \sim p_{X_v}, X_l \sim p_{X_l} \qquad X_a, X_v, X_l \sim p_{X_a, X_v, X_l}$$

$$T(\theta) : \mathscr{X}_a \times \mathscr{X}_v \times \mathscr{X}_l \to \mathbb{R}$$ **family of functions**

For MDM we extend Belghazi et al 2018 (MINE)

Training Losses

Training Loss $\mathscr{L} \triangleq \underbrace{\mathscr{L}_{down.}}_{\text{main task}} - \underbrace{\lambda \cdot \mathscr{L}_{MDM}}_{\text{mutual dependency term}}$

Notation $X_a \sim p_{X_a}, X_v \sim p_{X_v}, X_l \sim p_{X_l}$ $\quad X_a, X_v, X_l \sim p_{X_a,X_v,X_l}$

$T(\theta) : \mathscr{X}_a \times \mathscr{X}_v \times \mathscr{X}_l \to \mathbb{R}$ **family of functions**

For MDM we extend Belghazi et al 2018 (MINE)

	Formula
Wasserstein ($\mathcal{W}$)	$\mathbf{I}_{\mathcal{W}} \triangleq \sup_{\theta : T_\theta \in \mathbb{L}} \mathbb{E}_{p_{X_a X_v X_l}}[T_\theta] - \log\left[\mathbb{E}_{\prod_{j \in \{a,v,l\}} p_{X_j}}[T_\theta]\right].$
f-divergence (f)	$\mathbf{I}_f \triangleq \sup_{\theta} \mathbb{E}_{p_{X_a X_v X_l}}[T_\theta] - \mathbb{E}_{\prod_{j \in \{a,v,l\}} p_{X_j}}[e^{T_\theta - 1}].$
Kullback-Leibler (KL)	$\mathbf{I}_{kl} \triangleq \sup_{\theta} \mathbb{E}_{p_{X_a X_v X_l}}[T_\theta] - \log\left[\mathbb{E}_{\prod_{j \in \{a,v,l\}} p_{X_j}}[e^{T_\theta}]\right].$

Training Losses

Training Loss $\mathscr{L} \triangleq \underbrace{\mathscr{L}_{down.}}_{\text{main task}} - \underbrace{\lambda \cdot \mathscr{L}_{MDM}}_{\text{mutual dependency term}}$

Notation $X_a \sim p_{X_a}, X_v \sim p_{X_v}, X_l \sim p_{X_l}$ $\quad X_a, X_v, X_l \sim p_{X_a,X_v,X_l}$

$T(\theta) : \mathscr{X}_a \times \mathscr{X}_v \times \mathscr{X}_l \to \mathbb{R}$ **family of functions**

For MDM we extend Belghazi et al 2018 (MINE)

	Formula	Alternative Work
Wasserstein ($\mathcal{W}$)	$\mathbf{I}_{\mathcal{W}} \triangleq \sup_{\theta: T_\theta \in \mathbb{L}} \mathbb{E}_{p_{X_a X_v X_l}}[T_\theta] - \log\left[\mathbb{E}_{\prod_{j\in\{a,v,l\}} p_{X_j}}[T_\theta]\right].$	Ozair et al 2019.
f-divergence (f)	$\mathbf{I}_f \triangleq \sup_{\theta} \mathbb{E}_{p_{X_a X_v X_l}}[T_\theta] - \mathbb{E}_{\prod_{j\in\{a,v,l\}} p_{X_j}}[e^{T_\theta - 1}].$	KNIFE. Anonymous et al. 2022
Kullback-Leibler (KL)	$\mathbf{I}_{kl} \triangleq \sup_{\theta} \mathbb{E}_{p_{X_a X_v X_l}}[T_\theta] - \log\left[\mathbb{E}_{\prod_{j\in\{a,v,l\}} p_{X_j}}[e^{T_\theta}]\right].$	Oord et al 2018

Practical Implementation

Practical Implementation

Our method requiers to compute

$$MDM \triangleq \sup_{\theta} \quad \mathbb{E}_{p_{X_a X_v X_l}}[T_\theta] - g\left[\mathbb{E}_{\prod_{j \in \{a,v,l\}} p_{X_j}}[\,\cdot\,]\right].$$

Practical Implementation

Our method requiers to compute

$$MDM \triangleq \sup_{\theta} \quad \mathbb{E}_{p_{X_aX_vX_l}}[T_\theta] - g\left[\mathbb{E}_{\prod_{j\in\{a,v,l\}} p_{X_j}}[\,\cdot\,]\right].$$

Two Stage Procedure

Algorithm 1 Two-stage procedure to minimise multivariate dependency measures.

INPUT: $\mathcal{D}_n = \{(x_a^j, x_v^j, x_l^j), \forall j \in [1,n]\}$ multi-modal training dataset, m batch size, $\sigma_a, \sigma_v, \sigma_l : [1,m] \to [1,m]$ three permutations, θ_c weights of the deep classifier, θ weights of the statistical network T_θ.

Initialization: parameters $\boxed{\theta}$ and $\boxed{\theta_c}$

Build Negative Dataset:

$$\boxed{\bar{\mathcal{D}}_n = \{(x_a^{\sigma_a(j)}, x_v^{\sigma_v(j)}, x_l^{\sigma_l(j)}), \forall j \in [1,n]\}}$$

Optimization:

while (θ, θ_c) not converged **do**

  for $i \in [1, Unroll]$ **do**

    Sample from $\mathcal{D}_n$, $\mathcal{B} \sim p_{X_aX_vX_l}$

    Sample from $\bar{\mathcal{D}}_n$, $\bar{\mathcal{B}} \sim \prod_{j\in\{a,v,l\}} p_{X_j}$

    Update θ based on the empirical version of Eq. 6 or Eq. 7 or Eq. 8.

  end for

  Sample a batch $\mathcal{B}$ from $\mathcal{D}$

  Update θ_c with $\mathcal{B}$ using Eq. 5.

end while

OUTPUT: Classifiers weights θ_c

Practical Implementation

Our method requiers to compute

$$MDM \triangleq \boxed{\sup_{\theta}} \; \mathbb{E}_{p_{X_a X_v X_l}}[T_\theta] - g\left[\mathbb{E}_{\prod_{j\in\{a,v,l\}} p_{X_j}}[\,\cdot\,]\right].$$

Two Stage Procedure

1. Update θ to find the supremum

Algorithm 1 Two-stage procedure to minimise multivariate dependency measures.

INPUT: $\mathcal{D}_n = \{(x_a^j, x_v^j, x_l^j), \forall j \in [1,n]\}$ multi-modal training dataset, m batch size, $\sigma_a, \sigma_v, \sigma_l : [1,m] \to [1,m]$ three permutations, θ_c weights of the deep classifier, θ weights of the statistical network T_θ.

Initialization: parameters $\boxed{\theta}$ and $\boxed{\theta_c}$

Build Negative Dataset:

$$\boxed{\bar{\mathcal{D}}_n = \{(x_a^{\sigma_a(j)}, x_v^{\sigma_v(j)}, x_l^{\sigma_l(j)}), \forall j \in [1,n]\}}$$

Optimization:
- **while** (θ, θ_c) not converged **do**
  - **for** $i \in [1, Unroll]$ **do**
    - Sample from $\mathcal{D}_n$, $\mathcal{B} \sim p_{X_a X_v X_l}$
    - Sample from $\bar{\mathcal{D}}_n$, $\bar{\mathcal{B}} \sim \prod_{j\in\{a,v,l\}} p_{X_j}$
    - Update θ based on the empirical version of Eq. 6 or Eq. 7 or Eq. 8.
  - **end for**
  - Sample a batch $\mathcal{B}$ from $\mathcal{D}$
  - Update θ_c with $\mathcal{B}$ using Eq. 5.
- **end while**

OUTPUT: Classifiers weights θ_c

Practical Implementation

Our method requiers to compute

$$MDM \triangleq \boxed{\sup_{\theta}} \; \mathbb{E}_{p_{X_a X_v X_l}}[T_\theta] - g\left[\mathbb{E}_{\prod_{j\in\{a,v,l\}} p_{X_j}}[\,\cdot\,]\right].$$

Two Stage Procedure

1. Update θ to find the supremum
2. Update neural network using the total loss

Algorithm 1 Two-stage procedure to minimise multivariate dependency measures.

INPUT: $\mathcal{D}_n = \{(x_a^j, x_v^j, x_l^j), \forall j \in [1, n]\}$ multi-modal training dataset, m batch size, $\sigma_a, \sigma_v, \sigma_l : [1, m] \to [1, m]$ three permutations, θ_c weights of the deep classifier, θ weights of the statistical network T_θ.

Initialization: parameters $\boxed{\theta}$ and $\boxed{\theta_c}$

Build Negative Dataset:

$$\boxed{\bar{\mathcal{D}}_n = \{(x_a^{\sigma_a(j)}, x_v^{\sigma_v(j)}, x_l^{\sigma_l(j)}), \forall j \in [1, n]\}}$$

Optimization:
- **while** (θ, θ_c) not converged **do**
  - **for** $i \in [1, Unroll]$ **do**
    - Sample from $\mathcal{D}_n$, $\mathcal{B} \sim p_{X_a X_v X_l}$
    - Sample from $\bar{\mathcal{D}}_n$, $\bar{\mathcal{B}} \sim \prod_{j\in\{a,v,l\}} p_{X_j}$
    - Update θ based on the empirical version of Eq. 6 or Eq. 7 or Eq. 8.
  - **end for**
  - Sample a batch $\mathcal{B}$ from $\mathcal{D}$
  - Update θ_c with $\mathcal{B}$ using Eq. 5.
- **end while**

OUTPUT: Classifiers weights θ_c

Practical Implementation

Our method requiers to compute

$$MDM \triangleq \boxed{\sup_{\theta}} \; \mathbb{E}_{p_{X_a X_v X_l}}[T_\theta] - g\left[\mathbb{E}_{\prod_{j\in\{a,v,l\}} p_{X_j}}[\,\cdot\,]\right].$$

Two Stage Procedure

1. Update θ to find the supremum

2. Update neural network using the total loss

Algorithm 1 Two-stage procedure to minimise multivariate dependency measures.

INPUT: $\mathcal{D}_n = \{(x_a^j, x_v^j, x_l^j), \forall j \in [1, n]\}$ multi-modal training dataset, m batch size, $\sigma_a, \sigma_v, \sigma_l : [1, m] \to [1, m]$ three permutations, θ_c weights of the deep classifier, θ weights of the statistical network T_θ.

Initialization: parameters $\boxed{\theta}$ and $\boxed{\theta_c}$

Build Negative Dataset:

$$\boxed{\bar{\mathcal{D}}_n = \{(x_a^{\sigma_a(j)}, x_v^{\sigma_v(j)}, x_l^{\sigma_l(j)}), \forall j \in [1, n]\}}$$

Optimization:
 while (θ, θ_c) not converged **do**
  for $i \in [1, Unroll]$ **do**
   Sample from $\mathcal{D}_n$, $\mathcal{B} \sim p_{X_a X_v X_l}$
   Sample from $\bar{\mathcal{D}}_n$, $\bar{\mathcal{B}} \sim \prod_{j\in\{a,v,l\}} p_{X_j}$
   Update θ based on the empirical version of Eq. 6 or Eq. 7 or Eq. 8.
  end for
  Sample a batch $\mathcal{B}$ from $\mathcal{D}$
  Update θ_c with $\mathcal{B}$ using Eq. 5.
 end while

OUTPUT: Classifiers weights θ_c

Summary & Experimental Setting

Summary & Experimental Setting

Summary

$$\textbf{Total} \quad \underbrace{\mathscr{L}_{down.}}_{\textbf{main task}} - \underbrace{\lambda \cdot \mathscr{L}_{MDM}}_{\textbf{mutual dependency term}}$$

$$\mathscr{L}_{MDM} \triangleq MDM\left(p_{X_a X_v X_l}(x_a, x_v, x_l), \prod_{j \in \{a,v,l\}} p_{X_j}(x_j) \right)$$

Summary & Experimental Setting

Summary

Total $\underbrace{\mathscr{L}_{down.}}_{\text{main task}} - \underbrace{\lambda \cdot \mathscr{L}_{MDM}}_{\text{mutual dependency term}}$

$$\mathscr{L}_{MDM} \triangleq MDM\left(p_{X_aX_vX_l}(x_a, x_v, x_l), \prod_{j\in\{a,v,l\}} p_{X_j}(x_j)\right)$$

Require little intervention
+
can work on any models

Summary & Experimental Setting

Summary

Total $\underbrace{\mathscr{L}_{down.}}_{\text{main task}} - \underbrace{\lambda \cdot \mathscr{L}_{MDM}}_{\text{mutual dependency term}}$

$$\mathscr{L}_{MDM} \triangleq MDM\left(p_{X_aX_vX_l}(x_a, x_v, x_l), \prod_{j\in\{a,v,l\}} p_{X_j}(x_j)\right)$$

Require little intervention

+

can work on any models

Experimental Setting

- Sentiment Analysis
- Robustness Analysis
- Explainability using T_θ

Summary & Experimental Setting

Summary

Total $\underbrace{\mathscr{L}_{down.}}_{\text{main task}} - \underbrace{\lambda \cdot \mathscr{L}_{MDM}}_{\text{mutual dependency term}}$

$$\mathscr{L}_{MDM} \triangleq MDM\left(p_{X_aX_vX_l}(x_a, x_v, x_l), \prod_{j\in\{a,v,l\}} p_{X_j}(x_j)\right)$$

Require little intervention
+
can work on any models

Experimental Setting

- Sentiment Analysis
- Robustness Analysis
- Explainability using T_θ

Other results available in the paper

Overall Results

Overall Results

CMU-MOSEI CMU-MOSI

2,199/23,454 movie review videos

Sentiment Score in [-3,3]

Overall Results

CMU-MOSEI CMU-MOSI

2,199/23,454 movie review videos

Sentiment Score in [-3,3]

	Acc_7^h	Acc_2^h	MAE^l	$Corr^h$
CMU-MOSI				
$\mathcal{L}_\emptyset$	31.1	76.1	1.00	0.65
$\mathcal{L}_{kl}$	31.7	**76.4**	1.00	**0.66**
$\mathcal{L}_f$	**33.7**	76.2	1.02	**0.66**
$\mathcal{L}_\mathcal{W}$	33.5	**76.4**	**0.98**	**0.66**
CMU-MOSEI				
$\mathcal{L}_\emptyset$	44.2	75.0	0.72	0.52
$\mathcal{L}_{kl}$	44.5	75.6	0.70	0.53
$\mathcal{L}_f$	**45.5**	75.2	0.70	0.52
$\mathcal{L}_\mathcal{W}$	45.3	**75.9**	**0.68**	**0.54**

	CMU-MOSI				CMU-MOSEI			
	Acc_7^h	Acc_2^h	MAE^l	$Corr^h$	Acc_7^h	Acc_2^h	MAE^l	$Corr^h$
MFN								
$\mathcal{L}_\emptyset$	31.3	76.6	1.01	0.62	44.4	74.7	0.72	0.53
$\mathcal{L}_{kl}$	32.5	76.7	**0.96**	**0.65**	44.2	74.7	0.72	**0.57**
$\mathcal{L}_f$	35.7	77.4	**0.96**	**0.65**	46.1	**75.4**	**0.69**	0.56
$\mathcal{L}_\mathcal{W}$	**35.9**	**77.6**	**0.96**	**0.65**	**46.2**	75.1	**0.69**	0.56
LFN								
$\mathcal{L}_\emptyset$	31.9	76.9	1.00	0.63	45.2	74.2	0.70	0.54
$\mathcal{L}_{kl}$	32.6	**77.7**	0.97	0.63	46.1	75.3	0.68	**0.57**
$\mathcal{L}_f$	**35.6**	77.1	0.97	0.63	45.8	**75.4**	0.69	**0.57**
$\mathcal{L}_\mathcal{W}$	**35.6**	**77.7**	**0.96**	**0.67**	**46.2**	**75.4**	**0.67**	**0.57**
MAGBERT								
$\mathcal{L}_\emptyset$	40.2	84.7	0.79	0.80	46.8	84.9	**0.59**	0.77
$\mathcal{L}_{kl}$	**42.0**	**85.6**	**0.76**	**0.82**	47.1	85.4	**0.59**	**0.79**
$\mathcal{L}_f$	41.7	**85.6**	0.78	**0.82**	46.9	**85.6**	**0.59**	**0.79**
$\mathcal{L}_\mathcal{W}$	41.8	85.3	**0.76**	**0.82**	**47.8**	85.5	**0.59**	**0.79**
MAGXLNET								
$\mathcal{L}_\emptyset$	43.0	86.2	0.76	**0.82**	46.7	84.4	**0.59**	0.79
$\mathcal{L}_{kl}$	**44.5**	86.1	**0.74**	**0.82**	47.5	85.4	**0.59**	0.81
$\mathcal{L}_f$	43.9	86.6	**0.74**	**0.82**	47.4	85.0	**0.59**	0.81
$\mathcal{L}_\mathcal{W}$	44.4	**86.9**	**0.74**	**0.82**	**47.9**	**85.8**	**0.59**	**0.82**

Overall Results

CMU-MOSEI CMU-MOSI

2,199/23,454 movie review videos

Sentiment Score in [-3,3]

Simple fusion mechanism

	Acc_7^h	Acc_2^h	MAE^l	$Corr^h$
CMU−MOSI				
$\mathcal{L}_\emptyset$	31.1	76.1	1.00	0.65
$\mathcal{L}_{kl}$	31.7	**76.4**	1.00	**0.66**
$\mathcal{L}_f$	**33.7**	76.2	1.02	**0.66**
$\mathcal{L}_\mathcal{W}$	33.5	**76.4**	**0.98**	**0.66**
CMU−MOSEI				
$\mathcal{L}_\emptyset$	44.2	75.0	0.72	0.52
$\mathcal{L}_{kl}$	44.5	75.6	0.70	0.53
$\mathcal{L}_f$	**45.5**	75.2	0.70	0.52
$\mathcal{L}_\mathcal{W}$	45.3	**75.9**	**0.68**	**0.54**

Complex fusion mechanism

	CMU−MOSI				CMU−MOSEI			
	Acc_7^h	Acc_2^h	MAE^l	$Corr^h$	Acc_7^h	Acc_2^h	MAE^l	$Corr^h$
MFN								
$\mathcal{L}_\emptyset$	31.3	76.6	1.01	0.62	44.4	74.7	0.72	0.53
$\mathcal{L}_{kl}$	32.5	76.7	**0.96**	**0.65**	44.2	74.7	0.72	**0.57**
$\mathcal{L}_f$	35.7	77.4	**0.96**	**0.65**	46.1	**75.4**	**0.69**	0.56
$\mathcal{L}_\mathcal{W}$	**35.9**	**77.6**	**0.96**	**0.65**	**46.2**	75.1	**0.69**	0.56
LFN								
$\mathcal{L}_\emptyset$	31.9	76.9	1.00	0.63	45.2	74.2	0.70	0.54
$\mathcal{L}_{kl}$	32.6	**77.7**	0.97	0.63	46.1	75.3	0.68	**0.57**
$\mathcal{L}_f$	**35.6**	77.1	0.97	0.63	45.8	**75.4**	0.69	**0.57**
$\mathcal{L}_\mathcal{W}$	**35.6**	**77.7**	**0.96**	**0.67**	**46.2**	**75.4**	**0.67**	**0.57**
MAGBERT								
$\mathcal{L}_\emptyset$	40.2	84.7	0.79	0.80	46.8	84.9	**0.59**	0.77
$\mathcal{L}_{kl}$	**42.0**	**85.6**	**0.76**	**0.82**	47.1	85.4	**0.59**	**0.79**
$\mathcal{L}_f$	41.7	**85.6**	0.78	**0.82**	46.9	**85.6**	**0.59**	**0.79**
$\mathcal{L}_\mathcal{W}$	41.8	85.3	**0.76**	**0.82**	**47.8**	85.5	**0.59**	**0.79**
MAGXLNET								
$\mathcal{L}_\emptyset$	43.0	86.2	0.76	**0.82**	46.7	84.4	**0.59**	0.79
$\mathcal{L}_{kl}$	**44.5**	86.1	**0.74**	**0.82**	47.5	85.4	**0.59**	0.81
$\mathcal{L}_f$	43.9	86.6	**0.74**	**0.82**	47.4	85.0	**0.59**	0.81
$\mathcal{L}_\mathcal{W}$	44.4	**86.9**	**0.74**	**0.82**	**47.9**	**85.8**	**0.59**	**0.82**

Robustness to modality drop

Robustness to modality drop

1. Most of information carried by Text
    - Drop Text

Robustness to modality drop

1. Most of information carried by Text
   - Drop Text

How maximising the MDM affect robustness?

Robustness to modality drop

1. Most of information carried by Text
    - Drop Text

How maximising the MDM affect robustness?

Experiment

1. Train using 3 modalities
2. At inference we use only A, V or A+V

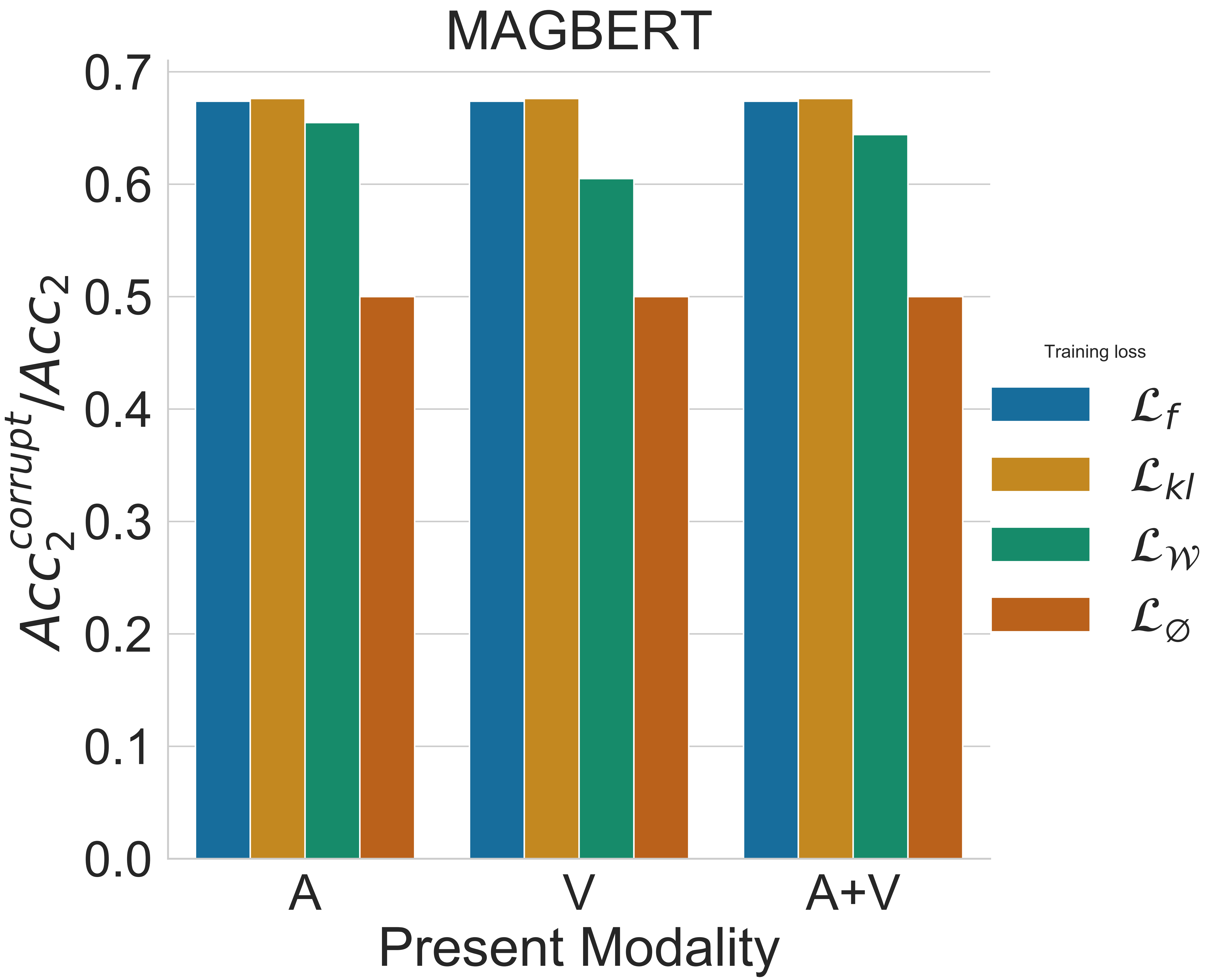
Robustness to modality drop

1. **Most of information carried by Text**
   - **Drop Text**

How maximising the MDM affect robustness?

Experiment

1. **Train using 3 modalities**
2. **At inference we use only A, V or A+V**



Robustness to modality drop

1. Most of information carried by Text
    - Drop Text

How maximising the MDM affect robustness?

Experiment

1. Train using 3 modalities
2. At inference we use only A, V or A+V

Our MDM loss improve robustness

Towards explainable representations

MDM

Towards explainable representations

Goal: Use low/high values or MDM to explain representations

MDM

Towards explainable representations

Goal: Use low/high values or MDM to explain representations

Spoken Transcripts	Acoustic and visual behaviour	MDM
um the story was all right	low energy monotonous voice + headshake	L

Towards explainable representations

Goal: Use low/high values or MDM to explain representations

Spoken Transcripts	Acoustic and visual behaviour	*MDM*
um the story was all right	low energy monotonous voice + headshake	L

Towards explainable representations

Goal: Use low/high values or MDM to explain representations

Spoken Transcripts	Acoustic and visual behaviour	*MDM*
um the story was all right	low energy monotonous voice + headshake	L
i mean its a Nicholas Sparks book it must be good	disappointed tone + neutral facial expression	L

Towards explainable representations

Goal: Use low/high values or MDM to explain representations

Spoken Transcripts	Acoustic and visual behaviour	*MDM*
um the story was all right	low energy monotonous voice + headshake	L
i mean its a Nicholas Sparks book it must be good	disappointed tone + neutral facial expression	L

Towards explainable representations

Goal: Use low/high values or MDM to explain representations

Spoken Transcripts	Acoustic and visual behaviour	*MDM*
um the story was all right	low energy monotonous voice + headshake	L
i mean its a Nicholas Sparks book it must be good	disappointed tone + neutral facial expression	L
the action is fucking awesome	head nod + excited voice	H

Towards explainable representations

Goal: Use low/high values or MDM to explain representations

Spoken Transcripts	Acoustic and visual behaviour	*MDM*
um the story was all right	low energy monotonous voice + headshake	L
i mean its a Nicholas Sparks book it must be good	disappointed tone + neutral facial expression	L
the action is fucking awesome	head nod + excited voice	H

Towards explainable representations

Goal: Use low/high values or MDM to explain representations

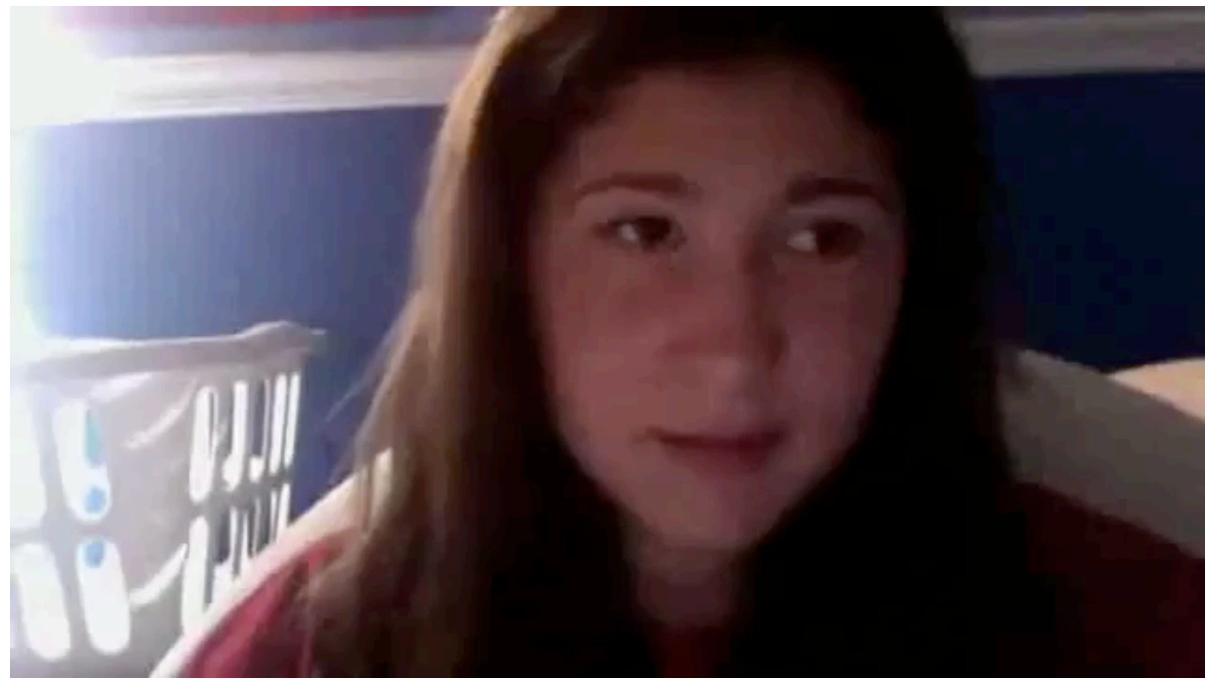
Spoken Transcripts	Acoustic and visual behaviour	MDM
um the story was all right	low energy monotonous voice + headshake	L
i mean its a Nicholas Sparks book it must be good	disappointed tone + neutral facial expression	L
the action is fucking awesome	head nod + excited voice	H
[l] it was cute you know the actors did a great job bringing the smurfs to life such as joe george lopez neil patrick harris katy perry and a fourth	multiple smiles	H

Towards explainable representations

Goal: Use low/high values or MDM to explain representations

Spoken Transcripts	Acoustic and visual behaviour	*MDM*
um the story was all right	low energy monotonous voice + headshake	L
i mean its a Nicholas Sparks book it must be good	disappointed tone + neutral facial expression	L
the action is fucking awesome	head nod + excited voice	H
[l] it was cute you know the actors did a great job bringing the smurfs to		
life such as joe george lopez neil patrick harris katy perry and a fourth	multiple smiles	H



Thanks for listening

Title: Improving Multimodal Fusion Via Mutual Dependency Maximisation

Link to Paper

Corresponding Authors:

Pierre Colombo