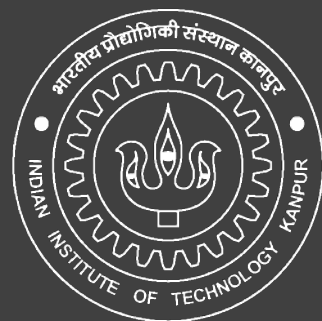# Ensemble Network for Ranking Images based on Visual Appeal
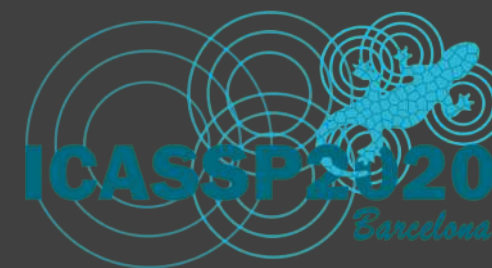
Sachin Singh[1], Victor Sanchez[2], Tanaya Guha[2]

presented by Sachin Singh

1 — INDIAN INSTITUTE OF TECHNOLOGY KANPUR

2 — WARWICK THE UNIVERSITY OF WARWICK

ICASSP2020 Barcelona

1

Outline

Outline

- Motivation

Outline

- Motivation
- Objective

Outline

- Motivation
- Objective
- Database Collection

Outline

- Motivation
- Objective
- Database Collection
- Proposed Methods/Models

Outline

- Motivation
- Objective
- Database Collection
- Proposed Methods/Models
- Results

Outline

- Motivation
- Objective
- Database Collection
- Proposed Methods/Models
- Results
- Conclusion

Motivation

Motivation

- With easy access to smartphones we are capturing and more and more pictures

Motivation

- With easy access to smartphones we are capturing and more and more pictures

- During social gatherings or casual group photoshoot we click many photos

Motivation

- With easy access to smartphones we are capturing and more and more pictures

- During social gatherings or casual group photoshoot we click many photos

- However we just like a few them, rest just fills our storage

How to Rank?

Objective

Objective

- To design a computational framework for ranking group photos

Objective

- To design a computational framework for ranking group photos
- Photos are assumed to be taken at the same event or within a short time span

Objective

- To design a computational framework for ranking group photos
- Photos are assumed to be taken at the same event or within a short time span
- Ranking is expected to match human perception of overall appeal of photos

Challenges

Challenges

- Subjective problem

Challenges

- Subjective problem
- Enormous complexity

Challenges

- Subjective problem
- Enormous complexity
- Unavailability of related dataset

Database Collection

Database Collection

- Collected around 300 group photos from web and volunteers

Database Collection

- Collected around 300 group photos from web and volunteers
- Photos were grouped into sets (each with 2-5 photos)

Database Collection

- Collected around 300 group photos from web and volunteers
- Photos were grouped into sets (each with 2-5 photos)
- Discarded sets which either have just 2 photos or its images are very similar

Database Collection

- Collected around 300 group photos from web and volunteers
- Photos were grouped into sets (each with 2-5 photos)
- Discarded sets which either have just 2 photos or its images are very similar
- Finally we have 70 sets of group images with 3 photos per set

Database Annotation

Database Annotation

- Build a project website and conducted an online survey

Database Annotation

- Build a project website and conducted an online survey
- Annotators were asked to provide the relative ranks of each photos within its own set

Database Annotation

- Build a project website and conducted an online survey

- Annotators were asked to provide the relative ranks of each photos within its own set

- Each photo is annotated by at least 5 annotators

Database Annotation

- Build a project website and conducted an online survey
- Annotators were asked to provide the relative ranks of each photos within its own set
- Each photo is annotated by at least 5 annotators
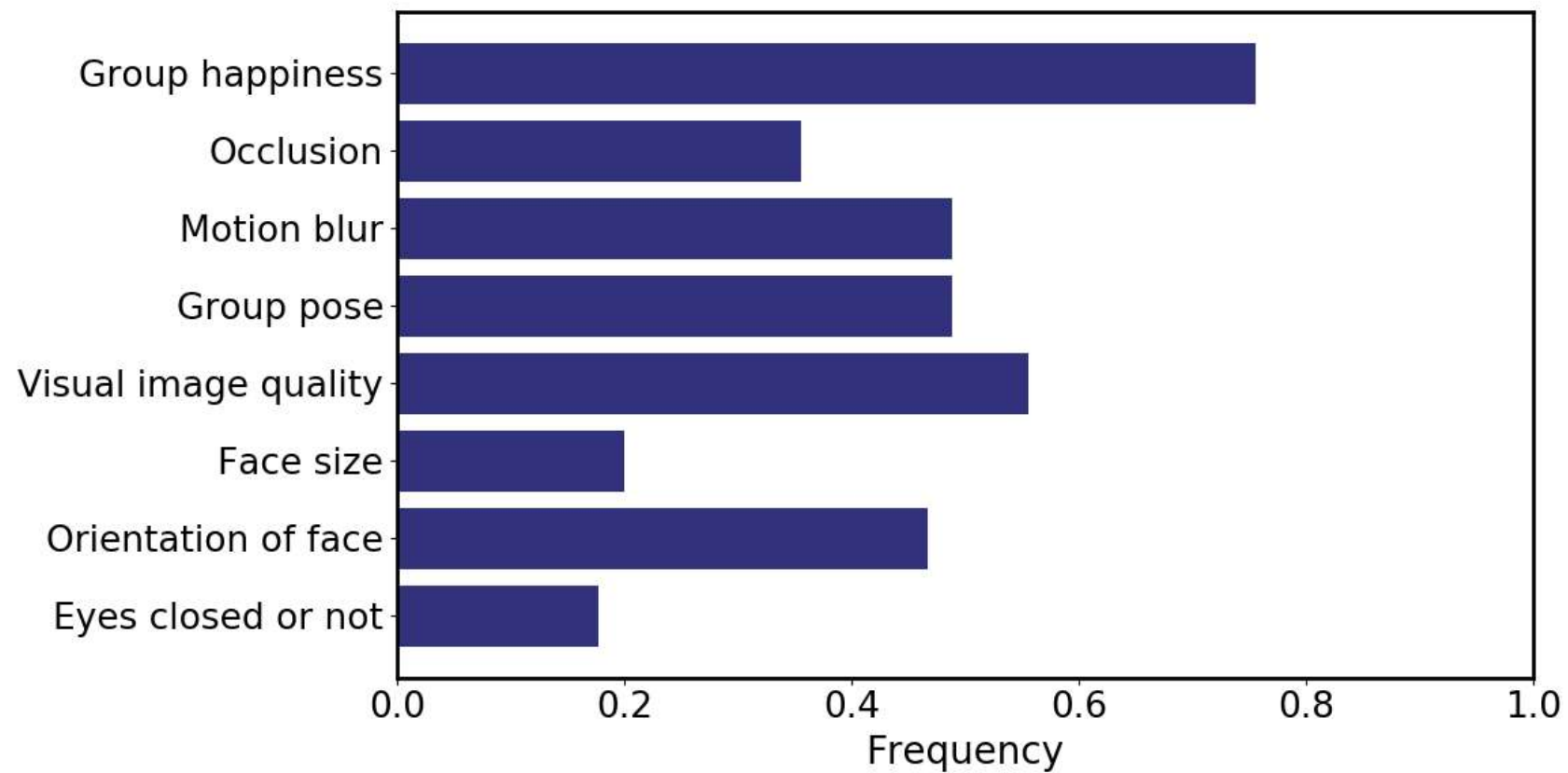- Final ranks were decided based on majority vote
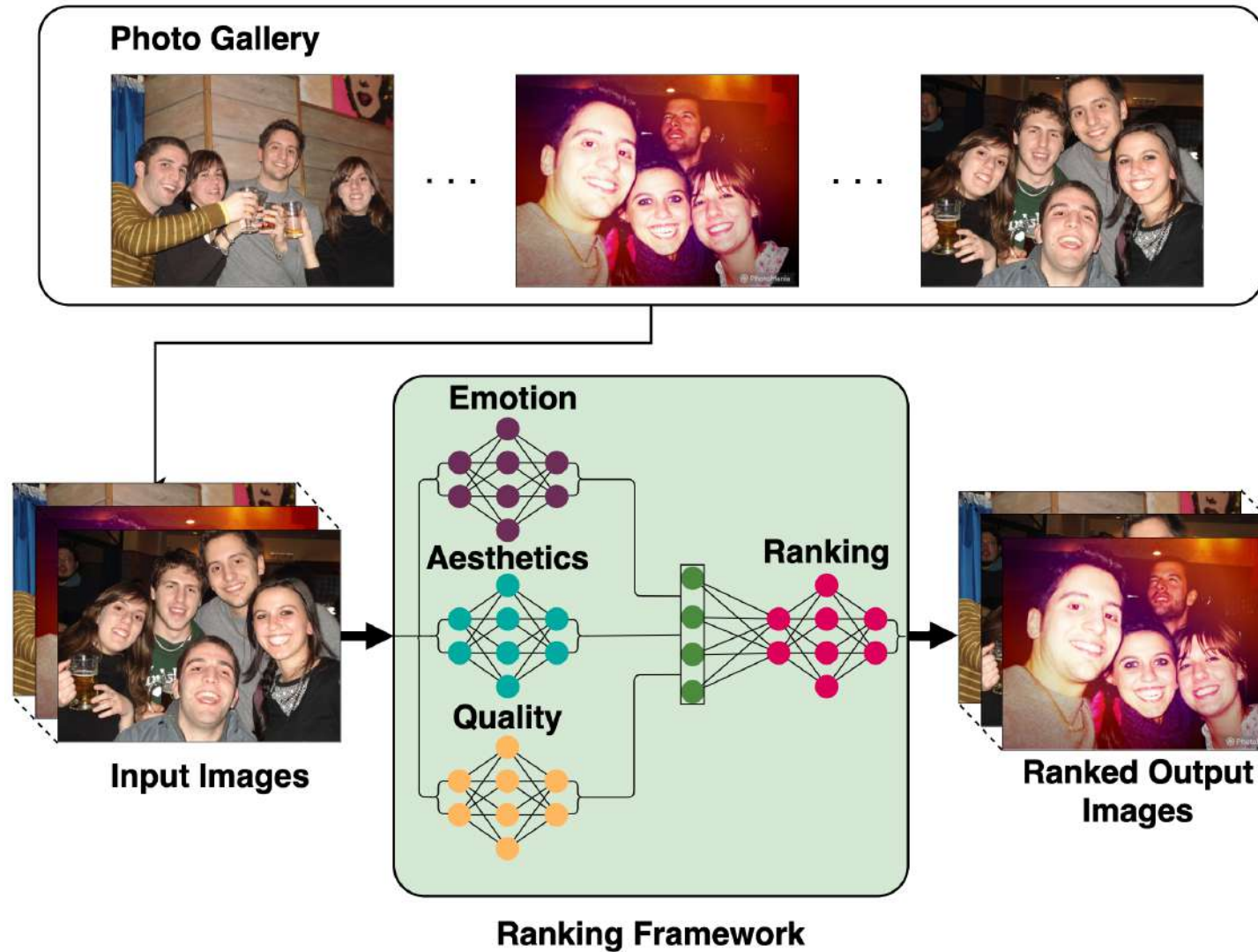
Project Website
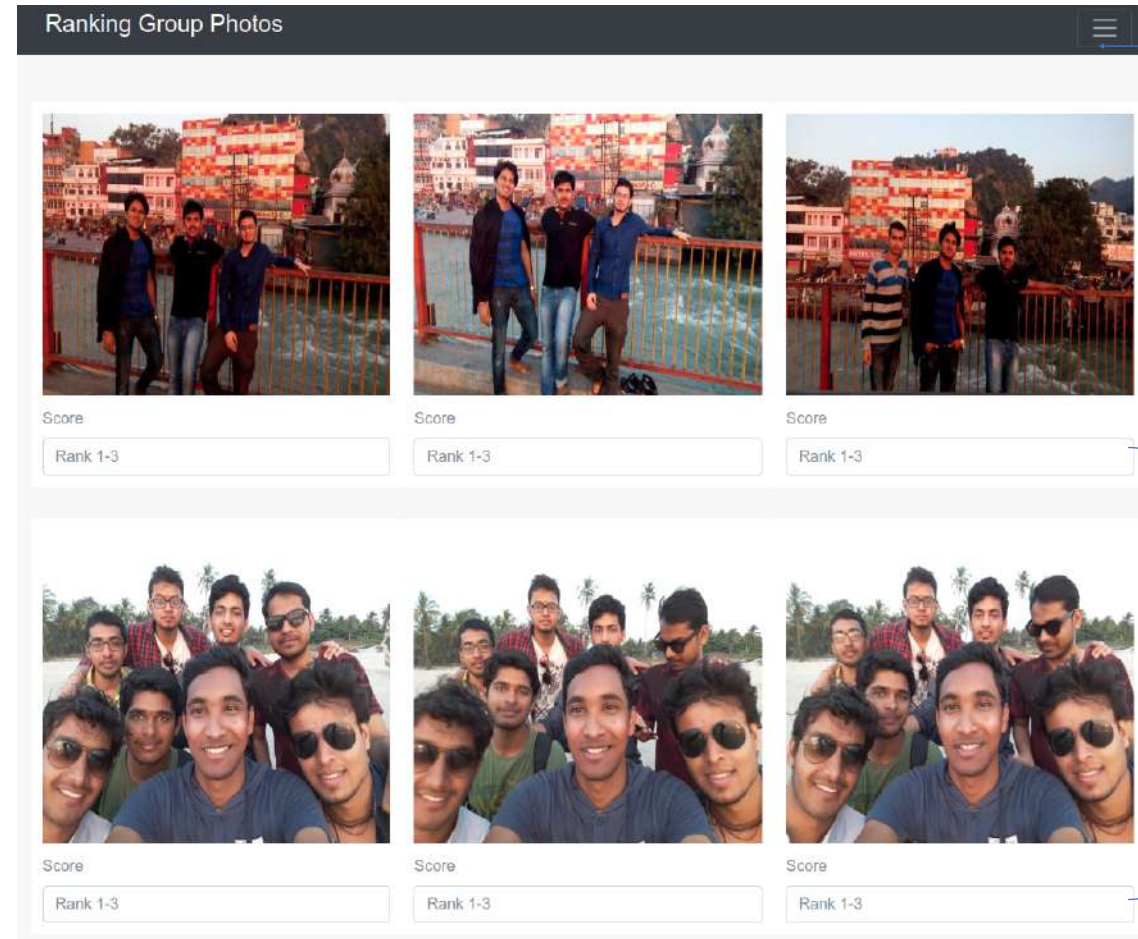
Instructions

Each row displaying one set

Space for relative ranks

Preliminary Results

Ranking Framework

Group Happiness

Group Happiness

- It's the overall happiness intensity expressed by a group photo

Image courtesy : https://www.istockphoto.com/in/video/slow-motion-beach-friends-group-selfie-gm467693556-61283828
http://jarilloherrero.mit.edu/photos/

Group Happiness

- It's the overall happiness intensity expressed by a group photo
- Includes both the global features e.g. context, group pose and local attributes like individual face expression

Image courtesy : https://www.istockphoto.com/in/video/slow-motion-beach-friends-group-selfie-gm467693556-61283828
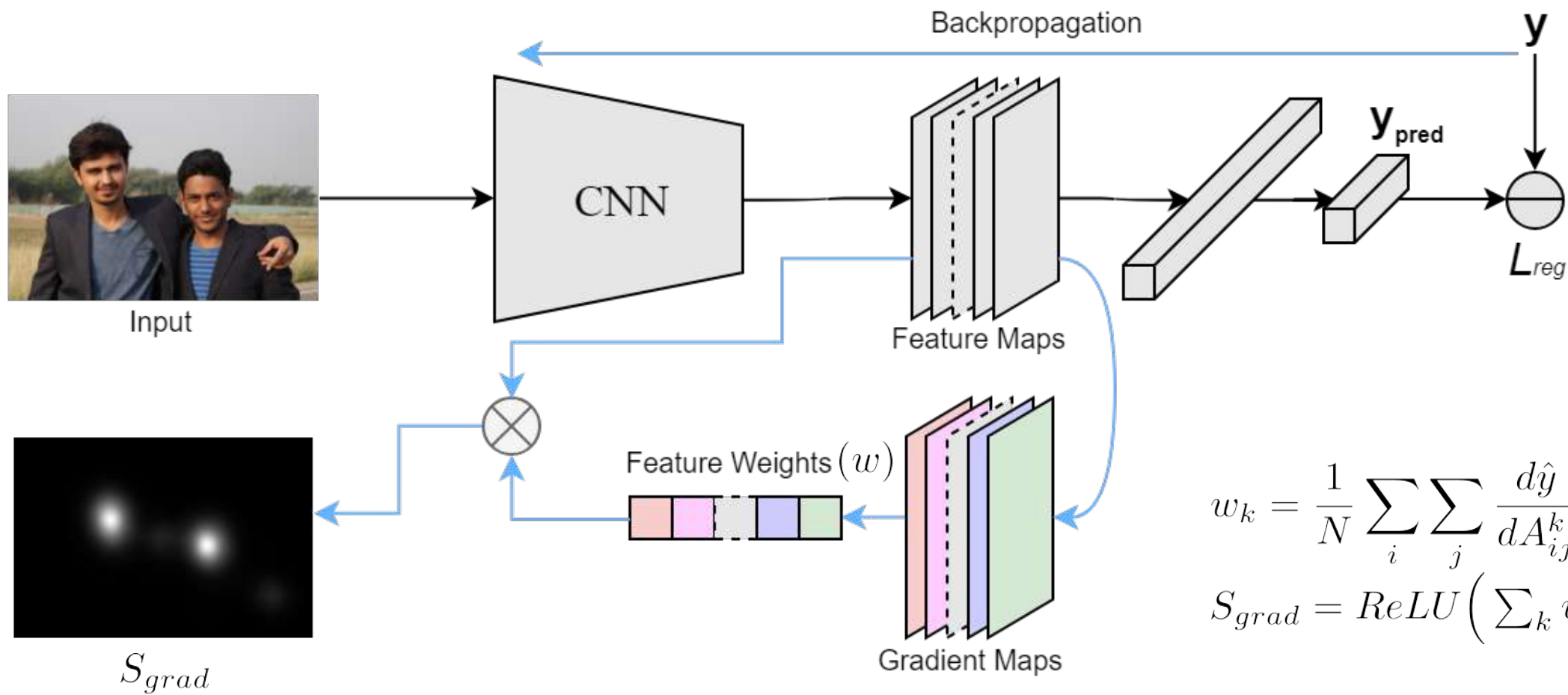http://jarilloherrero.mit.edu/photos/

13

HAPPEI Database

- Collected by Dhall et al. [2015]

- Composed of 2638 group images

- Images are labelled with six discrete labels (0-5) of happiness intensity for both individuals and group

neutral   small smile   big smile   small laugh   big laugh   thrilled
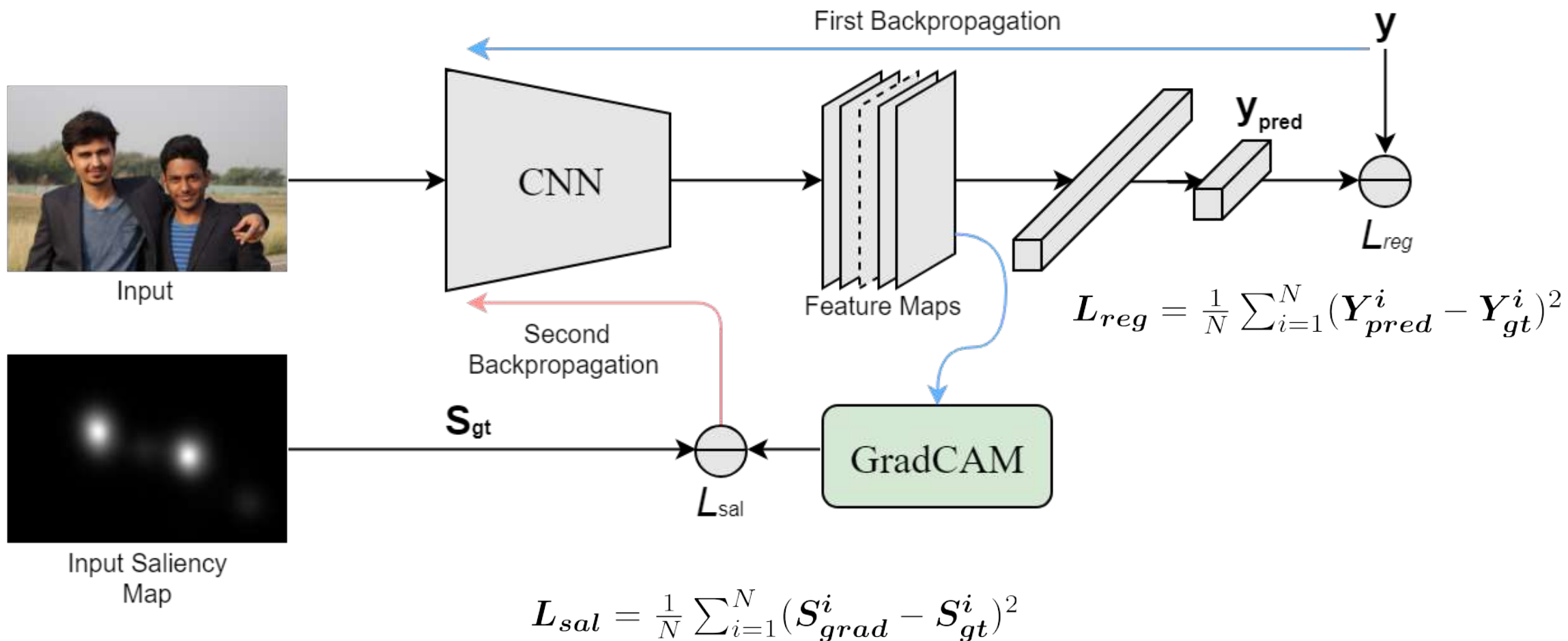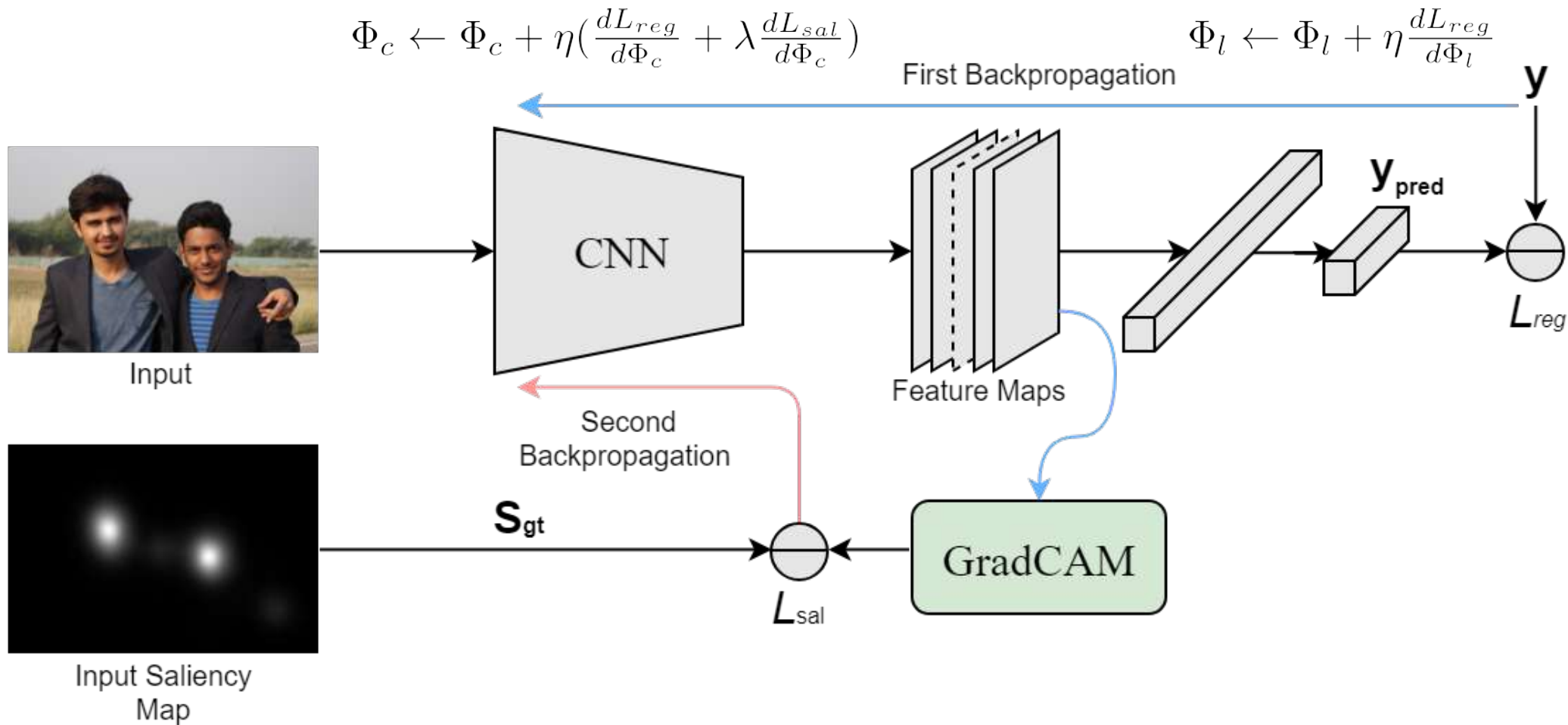
GradCAM (Silvaraju et al. [2017])

$$w_k = \frac{1}{N} \sum_i \sum_j \frac{d\hat{y}}{dA_{ij}^k}$$

$$S_{grad} = ReLU\left(\sum_k w_k A_k\right)$$

sCNN

First Backpropagation

CNN

Feature Maps

Input

$\mathbf{y}$

$\mathbf{y}_{pred}$

L_{reg}

$$L_{reg} = \frac{1}{N} \sum_{i=1}^{N} (Y_{pred}^i - Y_{gt}^i)^2$$

Second
Backpropagation

$\mathbf{S}_{gt}$

GradCAM

L_{sal}

Input Saliency
Map

$$L_{sal} = \frac{1}{N} \sum_{i=1}^{N} (S_{grad}^i - S_{gt}^i)^2$$

sCNN

$$\Phi_c \leftarrow \Phi_c + \eta\left(\frac{dL_{reg}}{d\Phi_c} + \lambda\frac{dL_{sal}}{d\Phi_c}\right) \qquad \Phi_l \leftarrow \Phi_l + \eta\frac{dL_{reg}}{d\Phi_l}$$

First Backpropagation

y

CNN

y$_{pred}$

Feature Maps

L_{reg}

Input

Second
Backpropagation

S$_{gt}$

L_{sal}

GradCAM

Input Saliency
Map

Attention Maps with and without Saliency loss

Without
Saliency loss

With
Saliency loss

Performance Evaluation

Table 1. Group happiness estimation results on HAPPEI database.

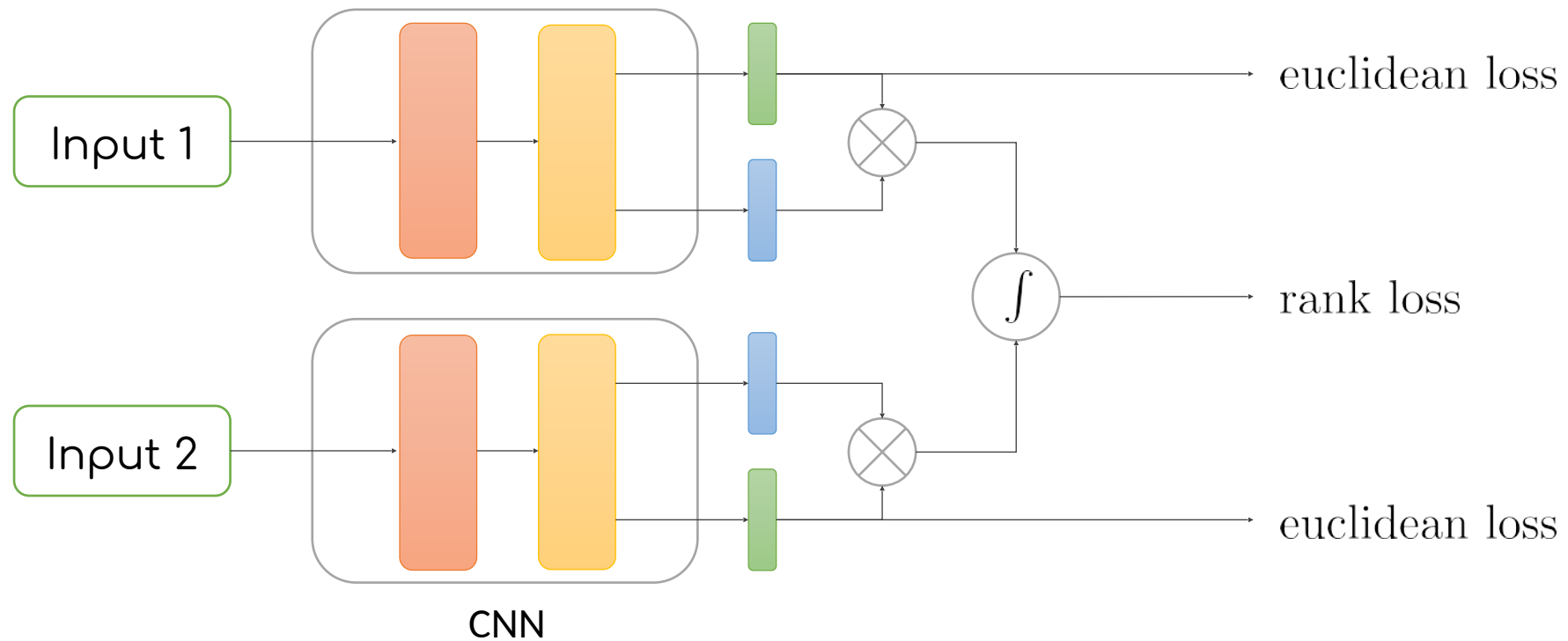| Method	MAE ↓
Mean emotion [18]	0.57
Dhall et al. [18]	**0.38**
Proposed without saliency	0.42
Proposed sCNN	**0.39**

Aesthetics
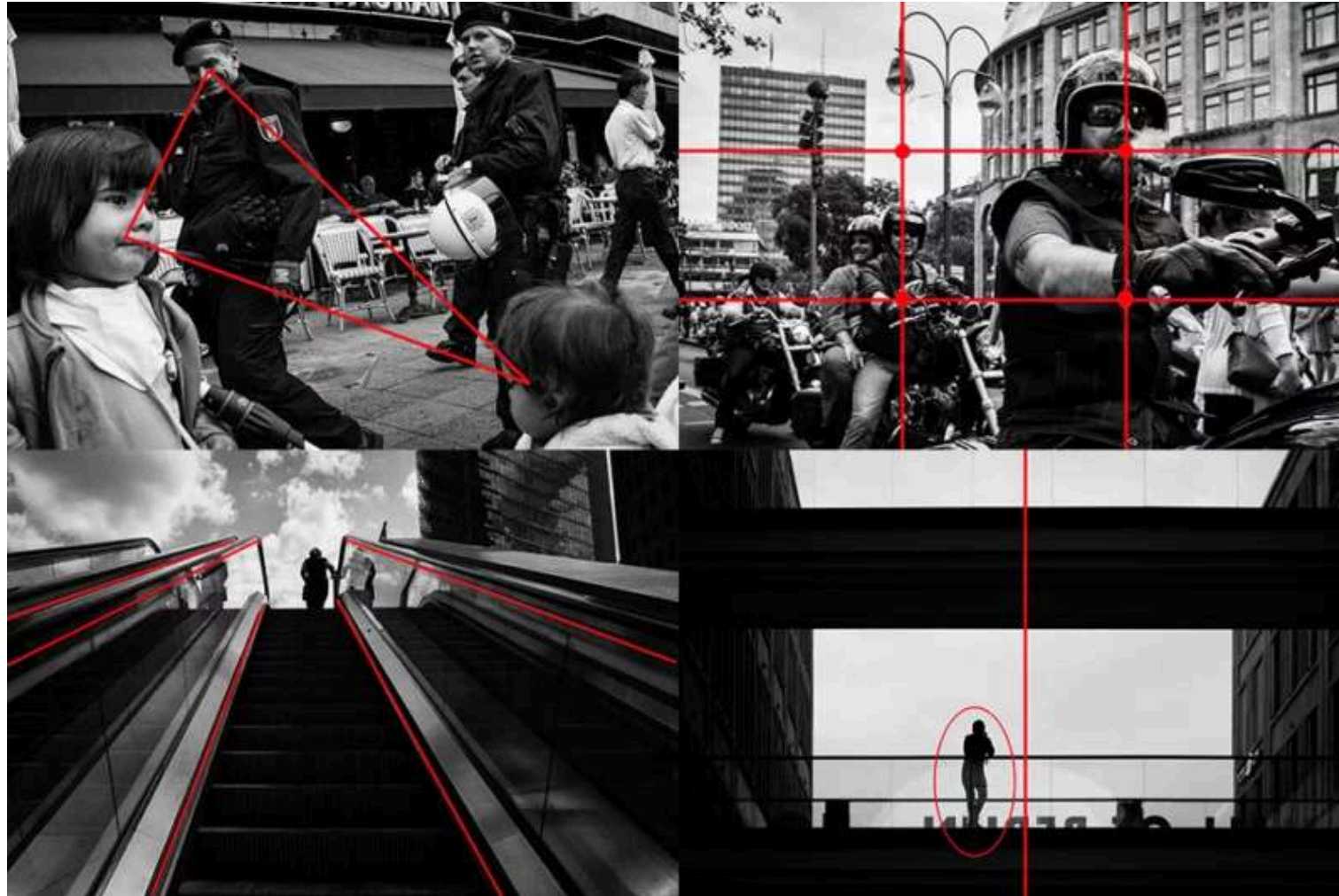
Aesthetics

Kong et al. [2016] proposed a CNN based model for recognising aesthetics in photos

Blind Image Quality (BIQ)

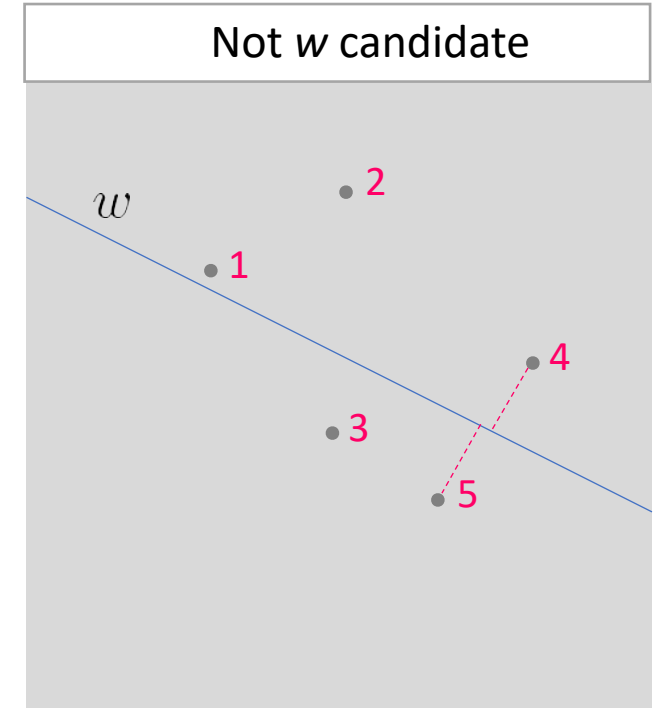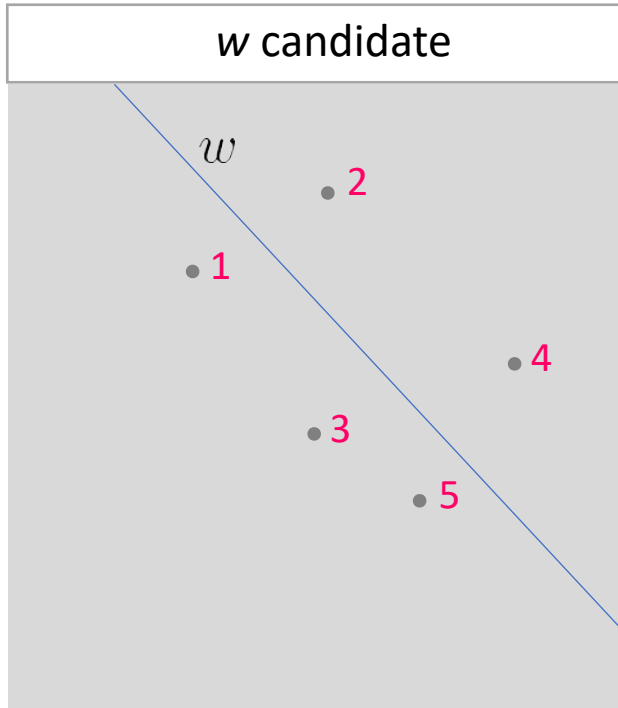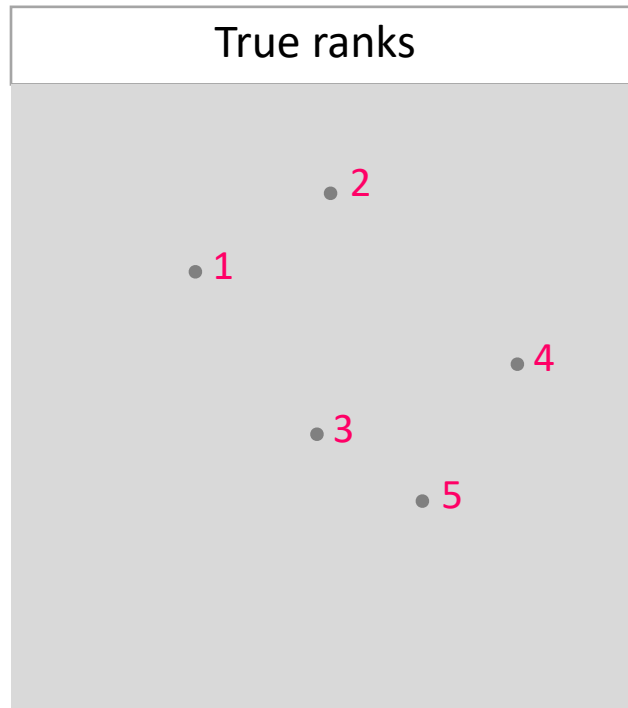- Image Quality assessment without any reference image

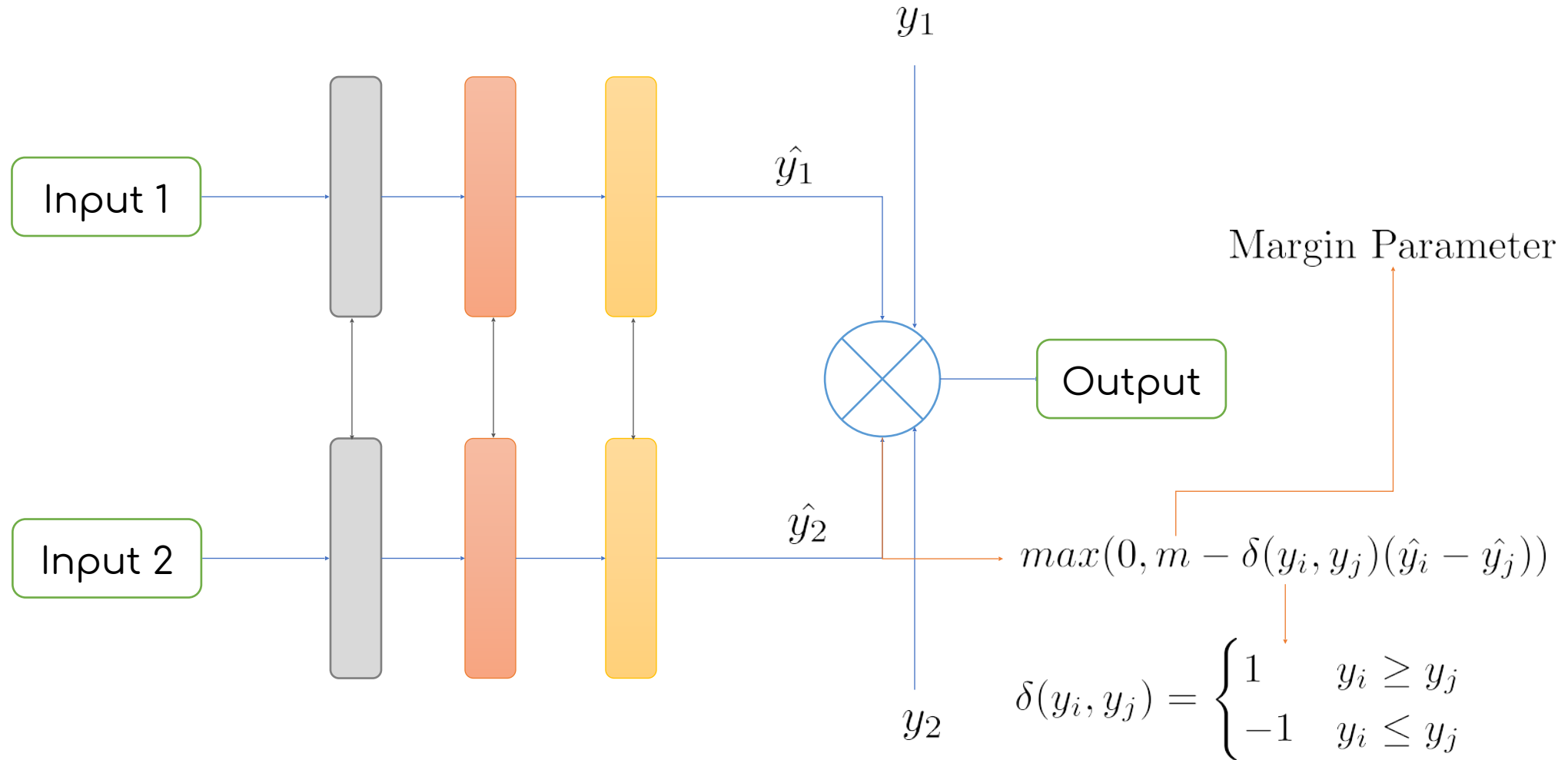
Score 4.7

Score 2.3

- In this work we used BRISQUE algorithm (Mittal et al. [2012]) for quantifying BIQ

Fusion : RankSVM

Fusion : RankNet

$$max(0, m - \delta(y_i, y_j)(\hat{y_i} - \hat{y_j}))$$

$$\delta(y_i, y_j) = \begin{cases} 1 & y_i \geq y_j \\ -1 & y_i \leq y_j \end{cases}$$

Margin Parameter

Evaluation Metrics

For comparing the performance of different ranking models we employed 3 evaluation metrics

- Spearman Ranking Correlation (higher the better)

- Best Image Match (BIM)

- Percentage of Swapped Pairs (PSP)

Evaluation Metrics : Proposed BIM

BIM – Best Image Match (higher the better)

$$BIM = \frac{\text{Total positive sets}}{\text{Total numbers of sets}} \times 100$$

positive set : set for which predicted best image matches with ground best image

Evaluation Metrics : Proposed PSP

PSP – Percentage of swapped pairs (lower the better)

$$PSP = \frac{\sum_i^N \text{Total no of swapped pairs in } i^{th} \text{ set}}{\sum_i^N \text{Total no of possible pairs in } i^{th} \text{ set}} \times 100$$

Here an image pair within a set is considered swapped if its predicted rank order is opposite to its ground truth rank order

Ranking Evaluations

Table 2. Group photo ranking performance on rGroup database.

Method	BIM ↑	PSP ↓	Corr (ρ) ↑
Avg. human performance	74.00	7.95	0.93
Individual channel			
Group happiness (sCNN)	27.14	39.70	0.21
Aesthetics [4]	37.10	27.80	0.52
Image quality	47.14	22.04	0.65
All channels			
Mean pooling	40.00	22.61	0.63
Max pooling	41.40	27.85	0.52
rankSVM	48.60	21.85	**0.69**
rankNet	**52.38**	**18.00**	**0.69**

Conclusion

Conclusion

- Collected a new database called Ranked Group Photos (rGroup) which is a collection of around 210 group photos taken during various social gatherings

Conclusion

- Collected a new database called Ranked Group Photos (rGroup) which is a collection of around 210 group photos taken during various social gatherings

- Proposed a CNN based architecture for predicting group happiness

Conclusion

- Collected a new database called Ranked Group Photos (rGroup) which is a collection of around 210 group photos taken during various social gatherings

- Proposed a CNN based architecture for predicting group happiness

- Proposed a multichannel computational framework for ranking a number of group photos taken at the same event within a short time span

Conclusion

- Collected a new database called Ranked Group Photos (rGroup) which is a collection of around 210 group photos taken during various social gatherings

- Proposed a CNN based architecture for predicting group happiness

- Proposed a multichannel computational framework for ranking a number of group photos taken at the same event within a short time span

- Defined two new evaluation metrics for evaluating ranking models – BIM and PSP

Conclusion

- Collected a new database called Ranked Group Photos (rGroup) which is a collection of around 210 group photos taken during various social gatherings

- Proposed a CNN based architecture for predicting group happiness

- Proposed a multichannel computational framework for ranking a number of group photos taken at the same event within a short time span

- Defined two new evaluation metrics for evaluating ranking models – BIM and PSP

- Need a larger dataset for training end-2-end model

Conclusion

- Collected a new database called Ranked Group Photos (rGroup) which is a collection of around 210 group photos taken during various social gatherings

- Proposed a CNN based architecture for predicting group happiness

- Proposed a multichannel computational framework for ranking a number of group photos taken at the same event within a short time span

- Defined two new evaluation metrics for evaluating ranking models – BIM and PSP

- Need a larger dataset for training end-2-end model

- Inclusion of other relevant visual cues in context of ranking group photos

Thank You