Seeing People in Social Context:

Recognizing People and Social Relationships

Gang Wang[1], Andrew Gallagher[2], Jiebo Luo[2],
and David Forsyth[1]

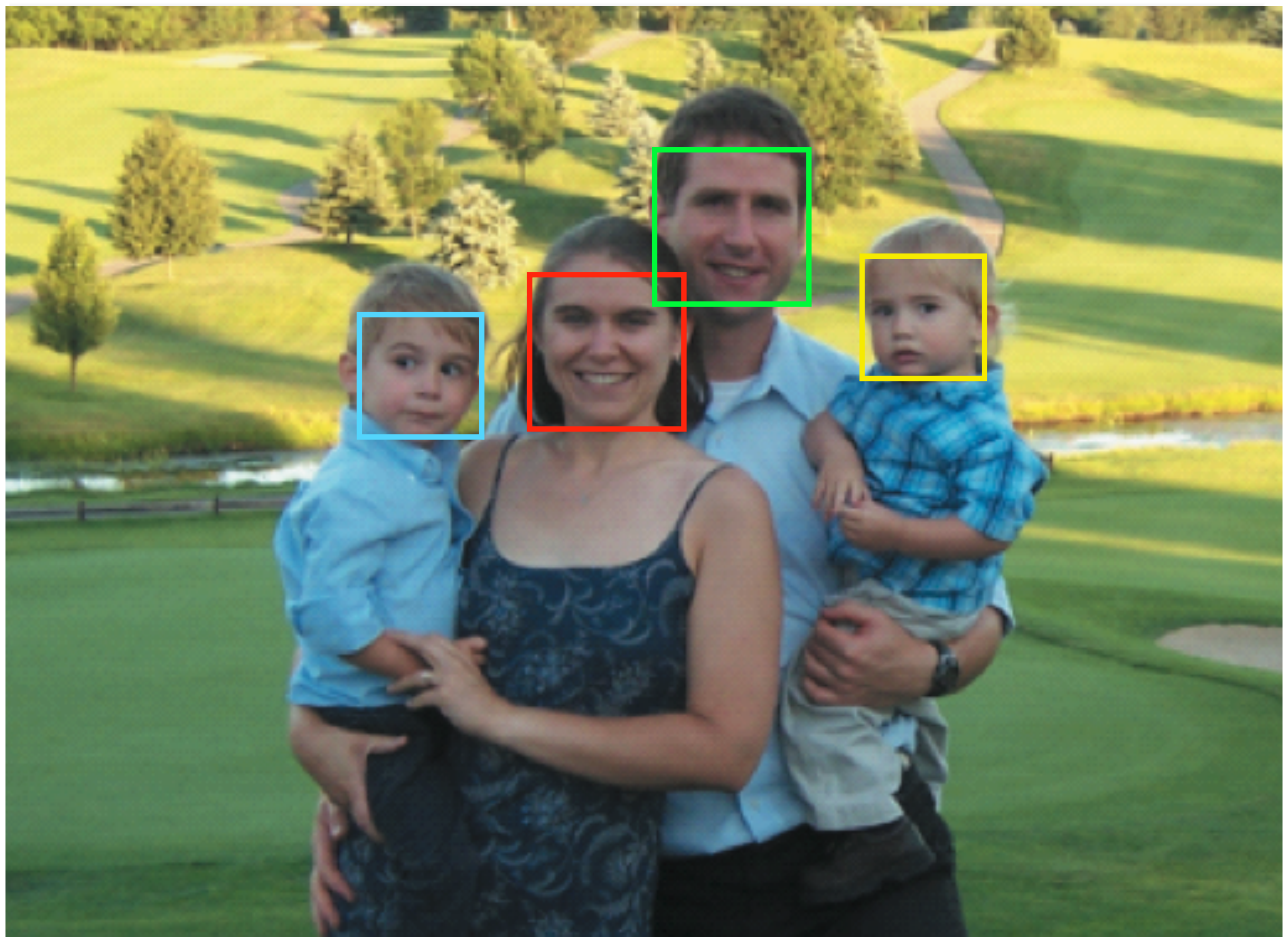
[1] University of Illinois at Urbana-Champaign, Urbana, IL
[2] Kodak Research Laboratories, Rochester, NY

ECCV 2010

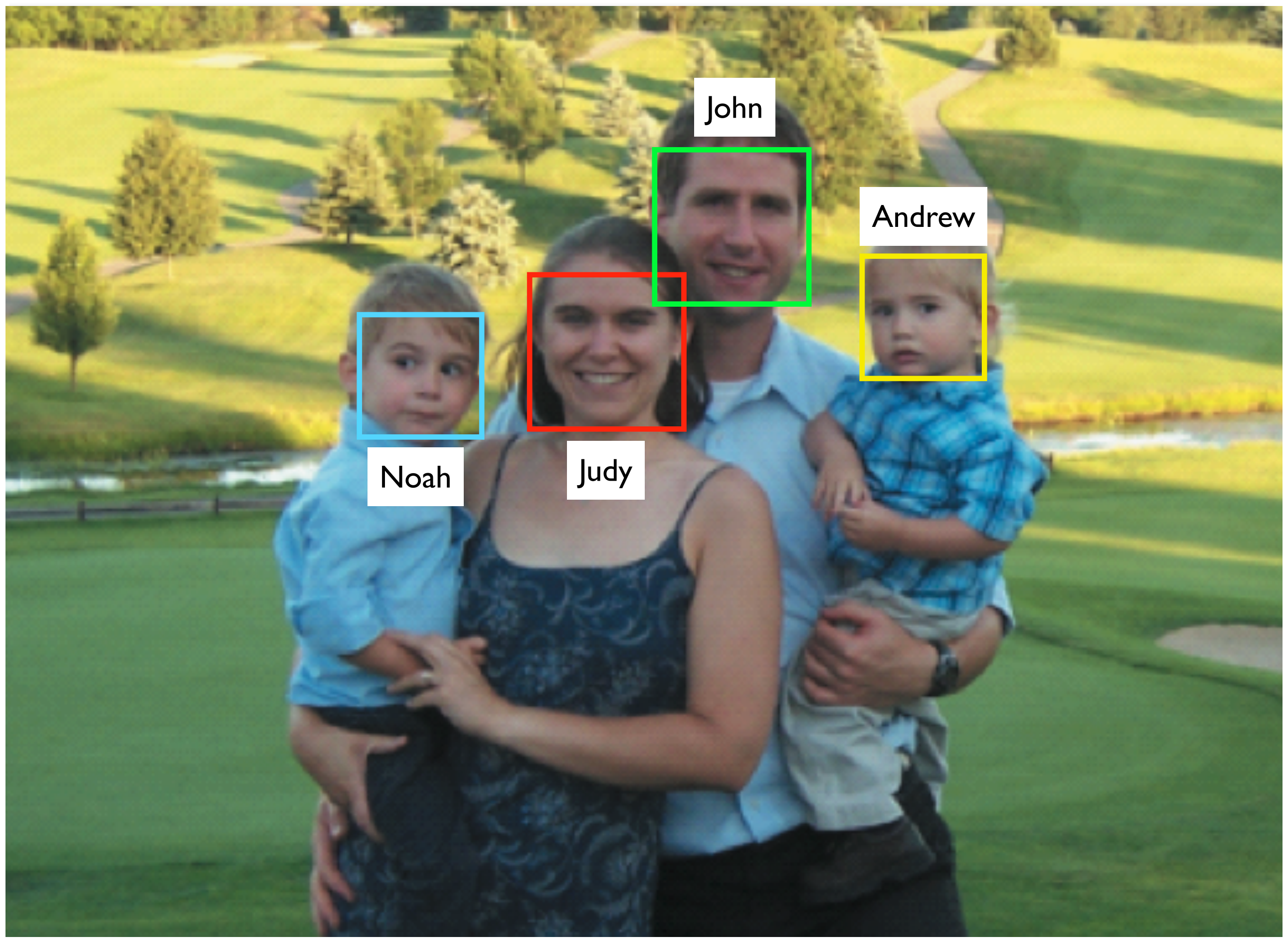
Presented by Aditya Sankar
CSE 590V

Traditional Approach

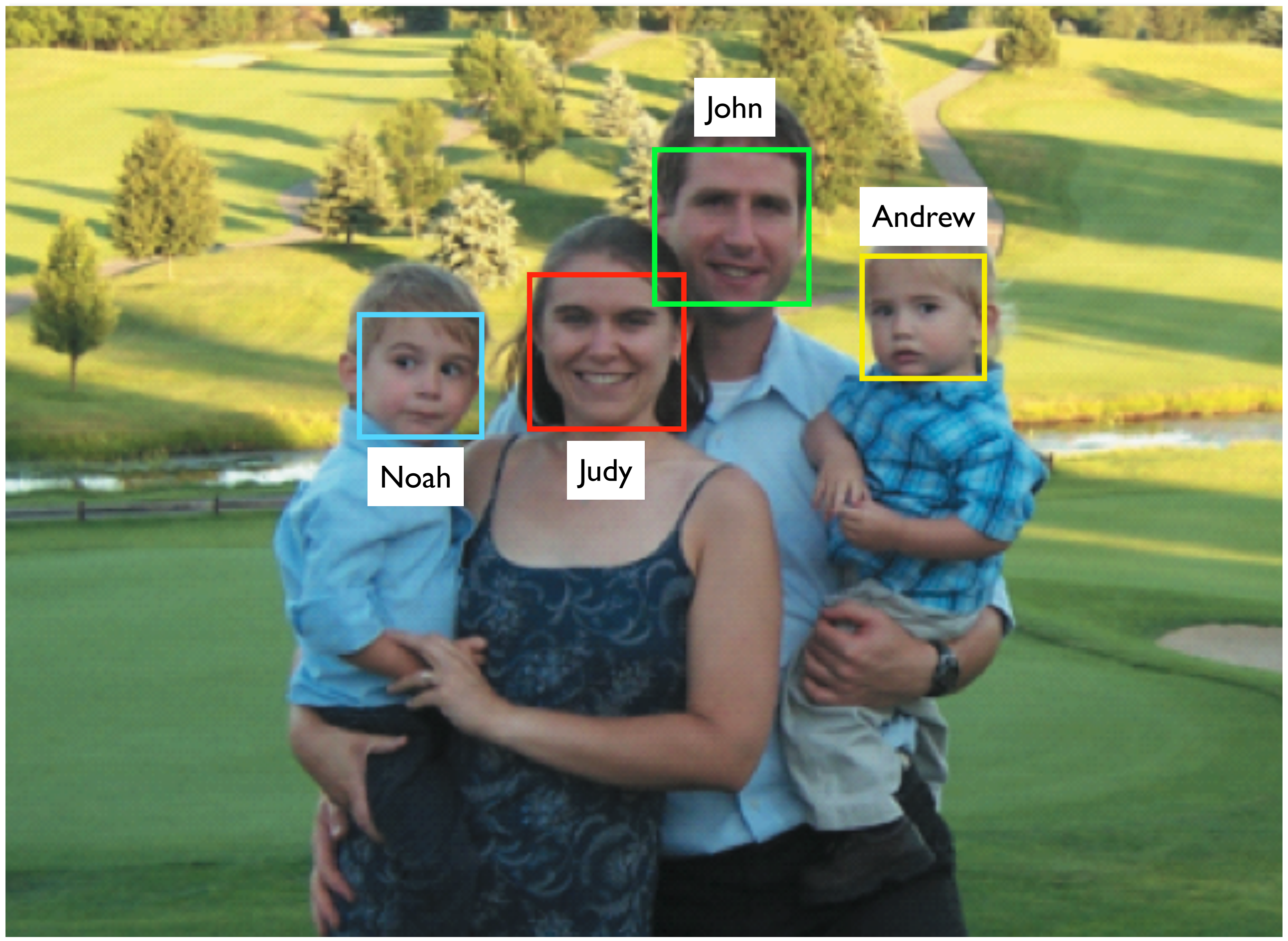
Traditional Approach



Traditional Approach

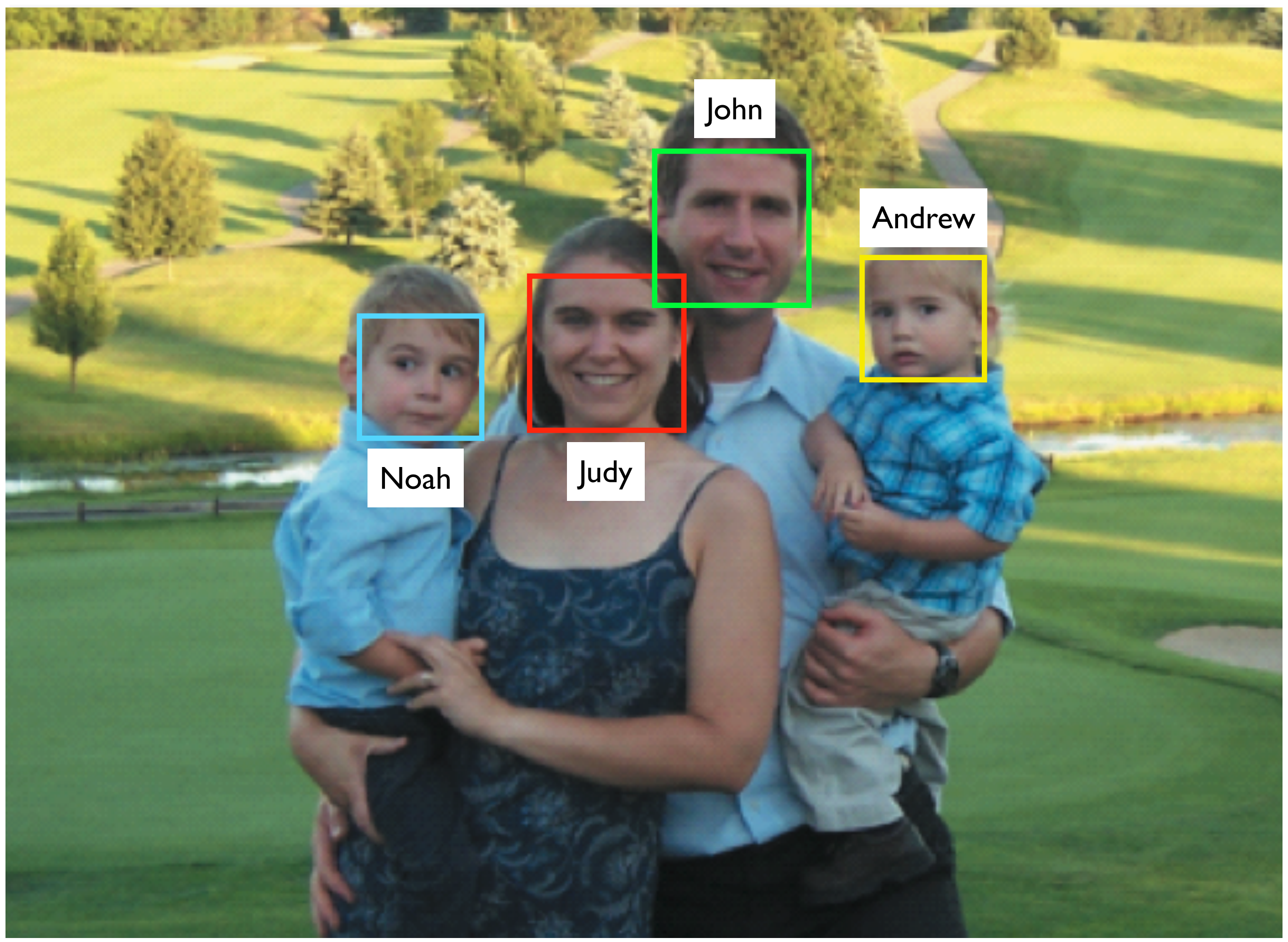


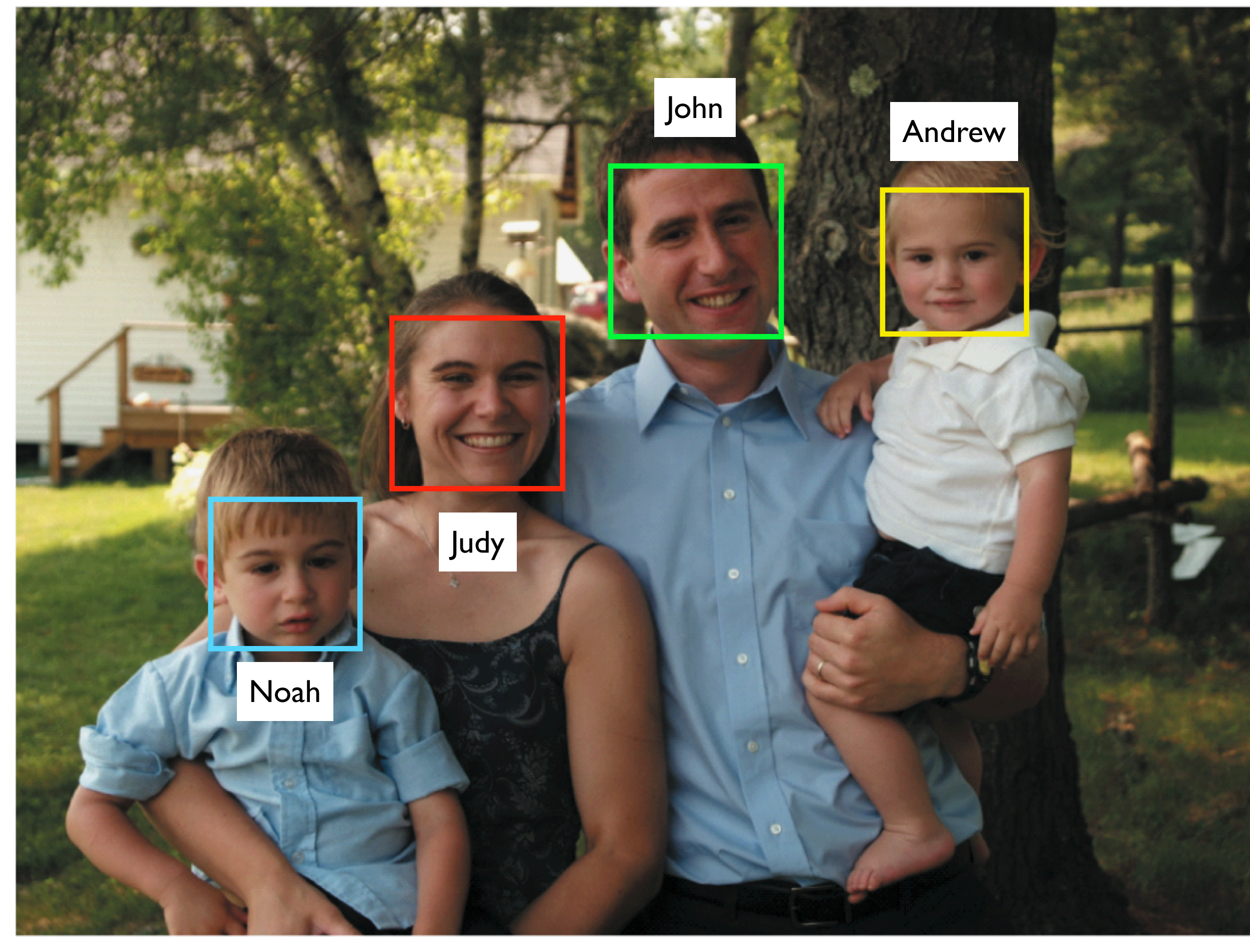
Traditional Approach



Construct Appearance Model

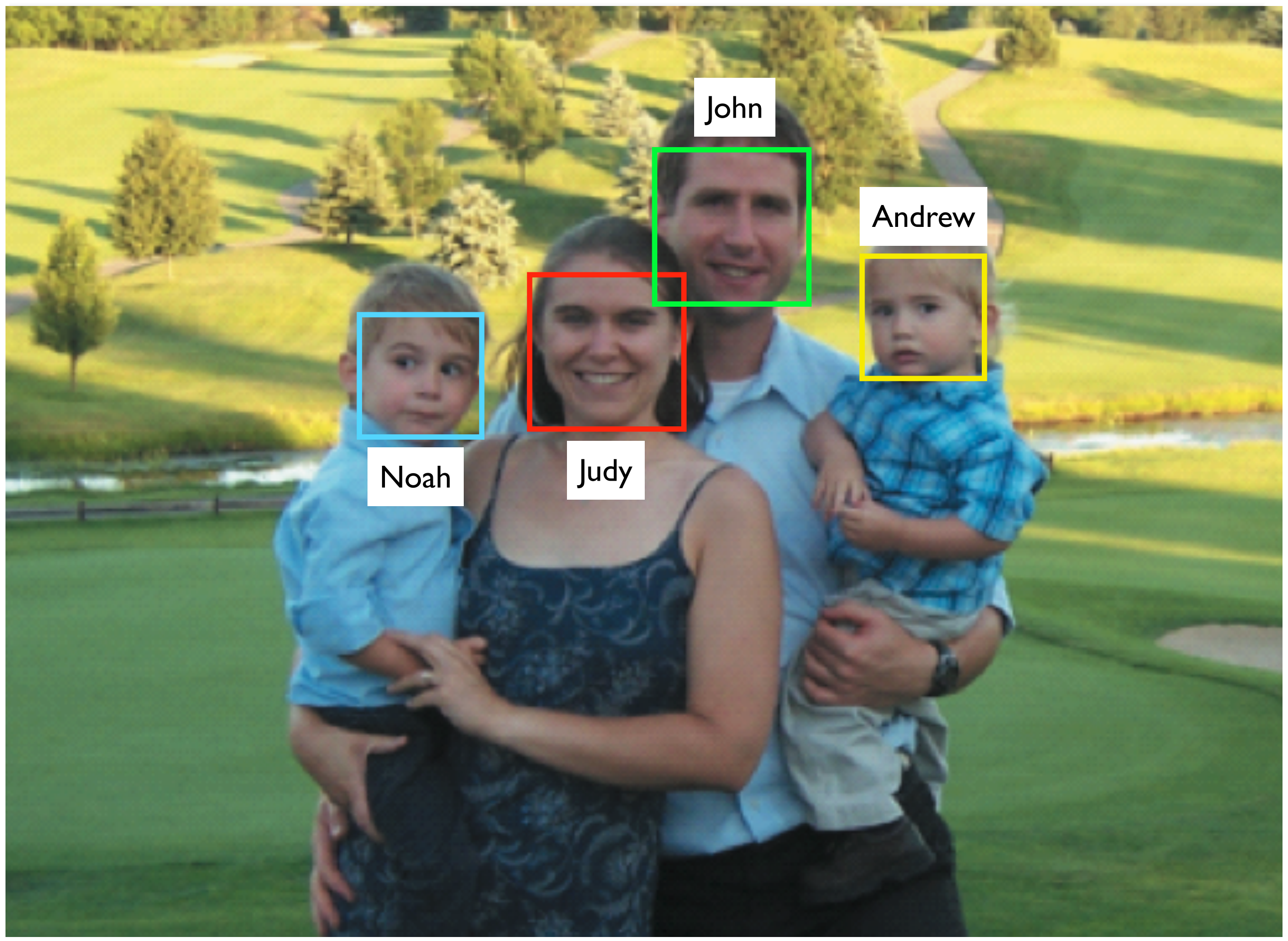
Traditional Approach

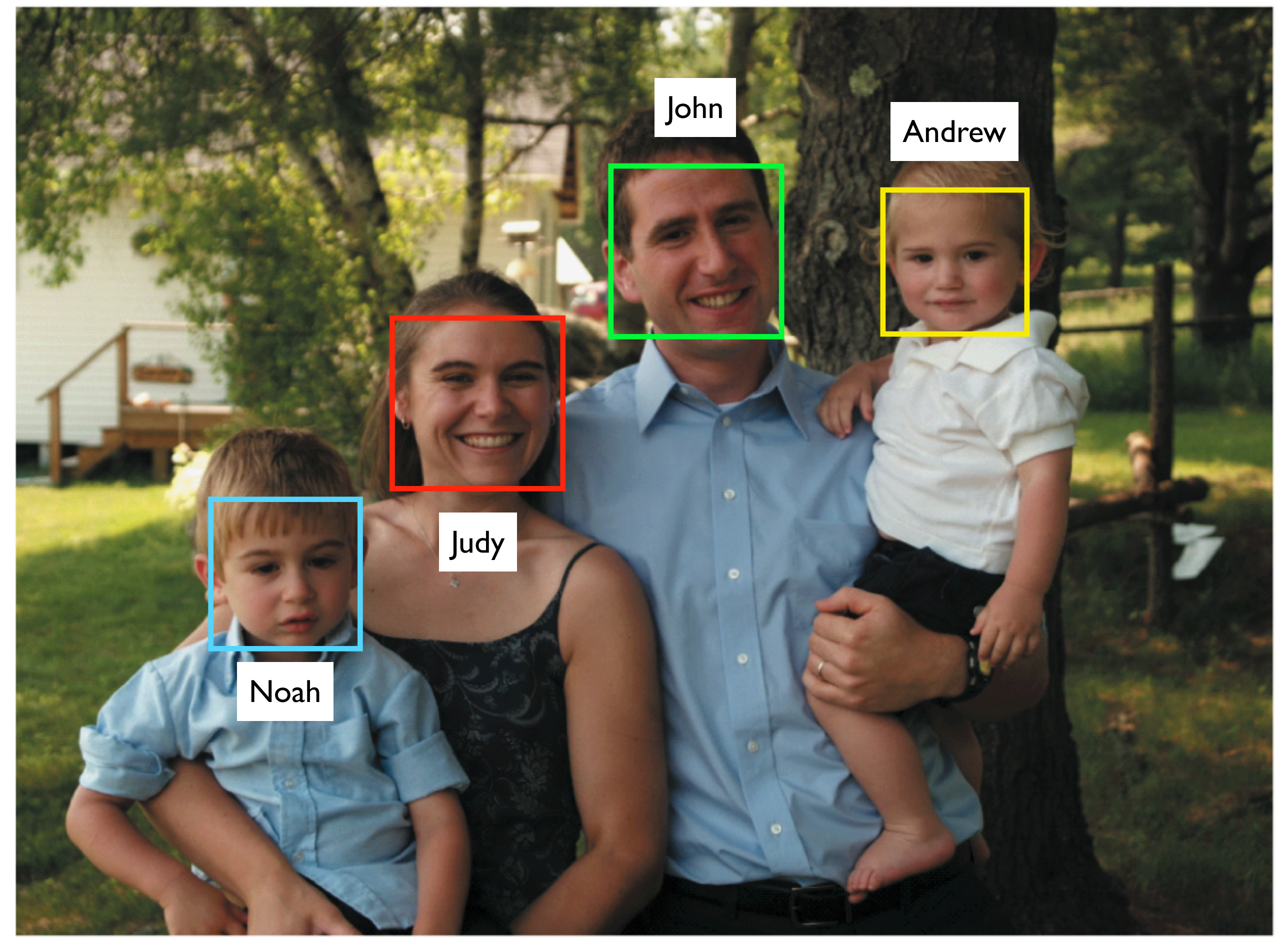




Construct Appearance Model ⟶ Recognize

Traditional Approach





Construct Appearance Model ⟶ Recognize

Does not work on weakly labeled data sets

Weak Labeling

Judy, John, Noah and Andrew in the UK

John, Judy and the kids at Eric's wedding

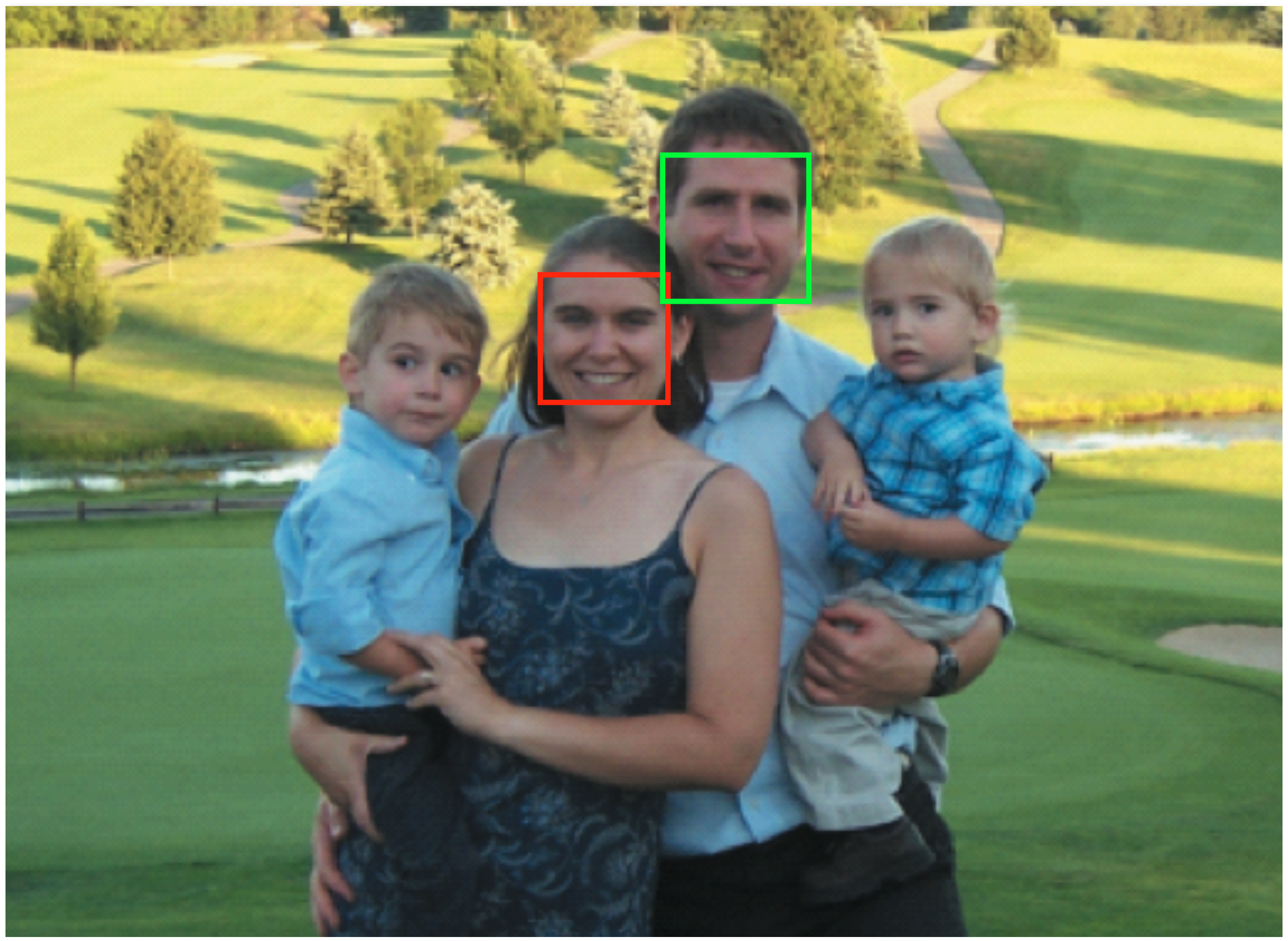
Photo albums, news captions, Flickr tags etc.

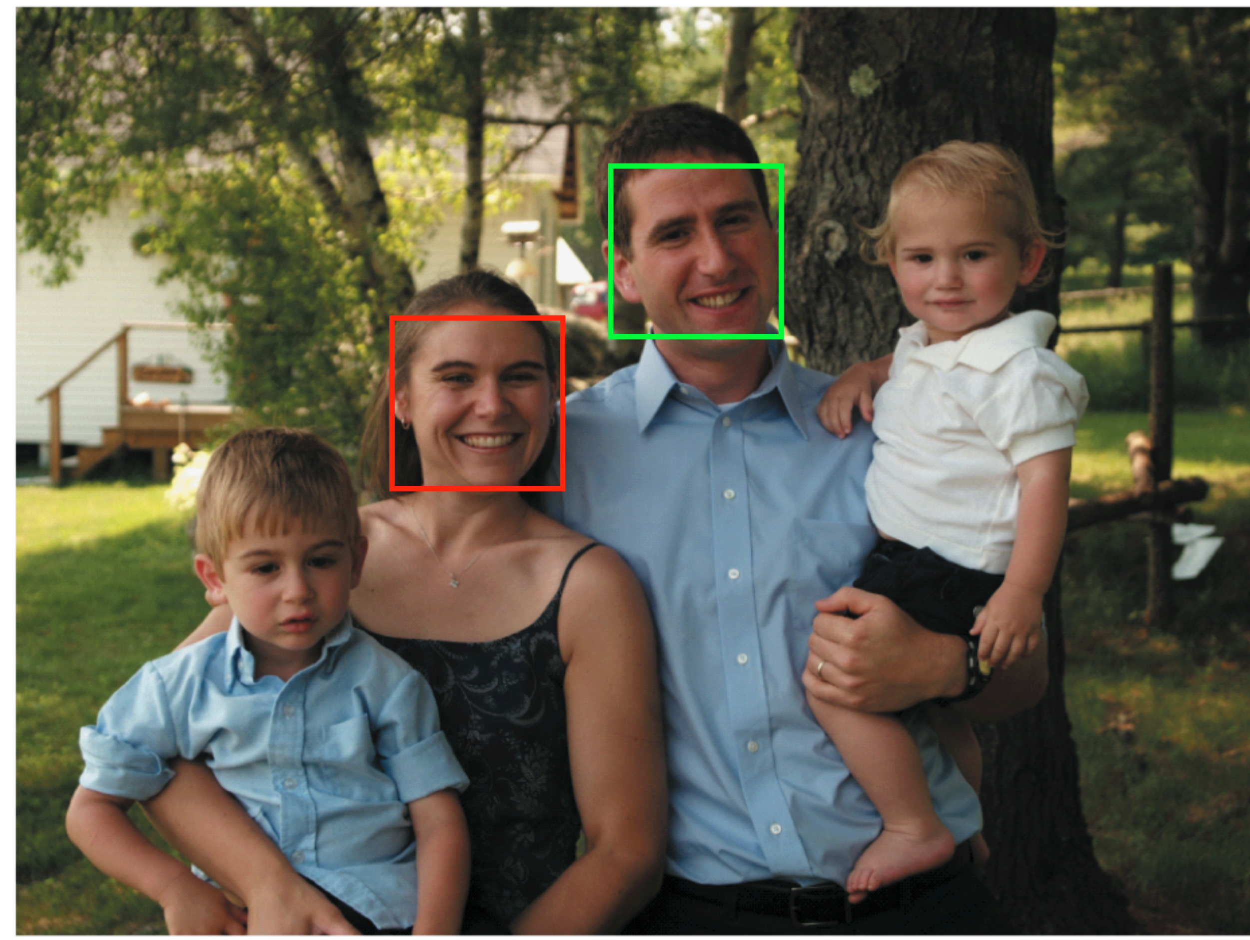
Label ambiguity increases learning difficulty

People in personal image collections are generally not strangers

Social relationships often exhibit certain visual patterns

People in personal image collections are generally not strangers





Social relationships often exhibit certain visual patterns

In this case:

- Husband and wife are in close proximity
- Husband is taller

Can we improve face recognition by considering these social relationships?

Training input:

Daisy, Noah

Edward, Daisy & Noah

Can we improve face recognition by considering these social relationships?

Training input:

Daisy, Noah

Edward, Daisy & Noah

Social relationships:

Daisy-Noah -> sibling
Daisy-Edward -> sibling
Noah-Edward -> sibling

Birth years:

Daisy: 2002
Noah: 2004
Edward: 2005

Can we improve face recognition by considering these social relationships?

Training input:

Daisy, Noah

Edward, Daisy & Noah

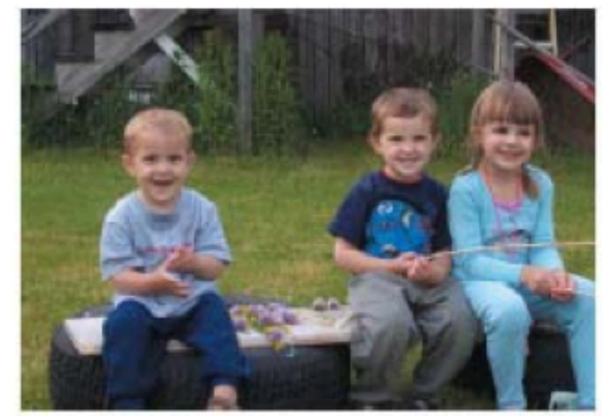
Social relationships:

Daisy-Noah -> sibling
Daisy-Edward -> sibling
Noah-Edward -> sibling

Birth years:

Daisy: 2002
Noah: 2004
Edward: 2005

Test:

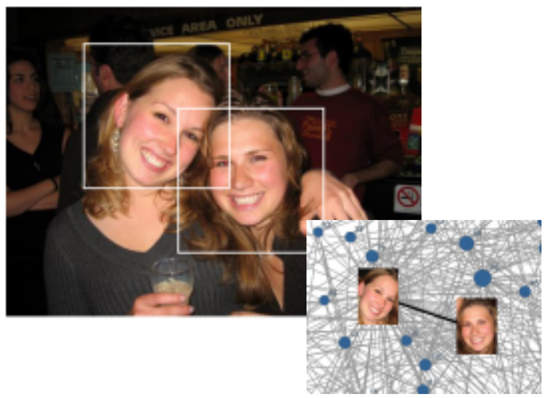


? ? ?

? ?

Related Work

Automatic Face Annotation



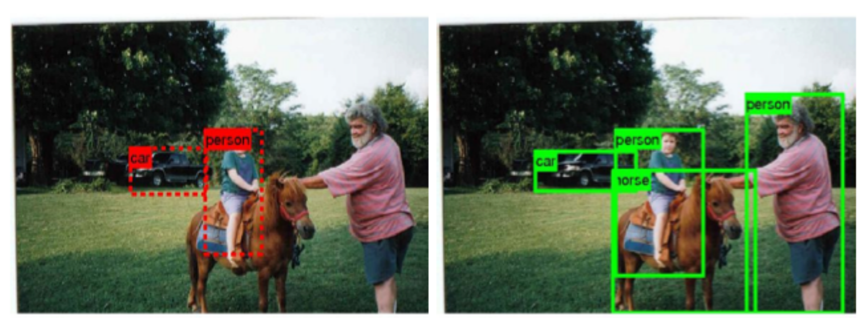
Stone et al. "Autotagging Facebook", CVPR 2008

Weakly Labeled Images

President George W. Bush makes a statement in the Rose Garden while Secretary of **Defense Donald Rumsfeld** looks on, July 23, 2003. Rumsfeld said the United States would release graphic photographs of the dead sons of **Saddam Hussein** to prove they were killed by American troops. Photo by Larry Downing/Reuters

Berg et al. "Names and Faces", CVPR 2004

Contextual Features



Divvala et al. "An Empirical Study of Context in Object Detection", CVPR 2008

Representing Social Relationships

r_{ij}: social relationship between i^{th} and j^{th} person

mother-child	father-child	grandparent-child	husband-wife	siblings
child-mother	child-father	child-grandparent	wife-husband	

f_{ij}: social relationship 'features' between i^{th} and j^{th} face

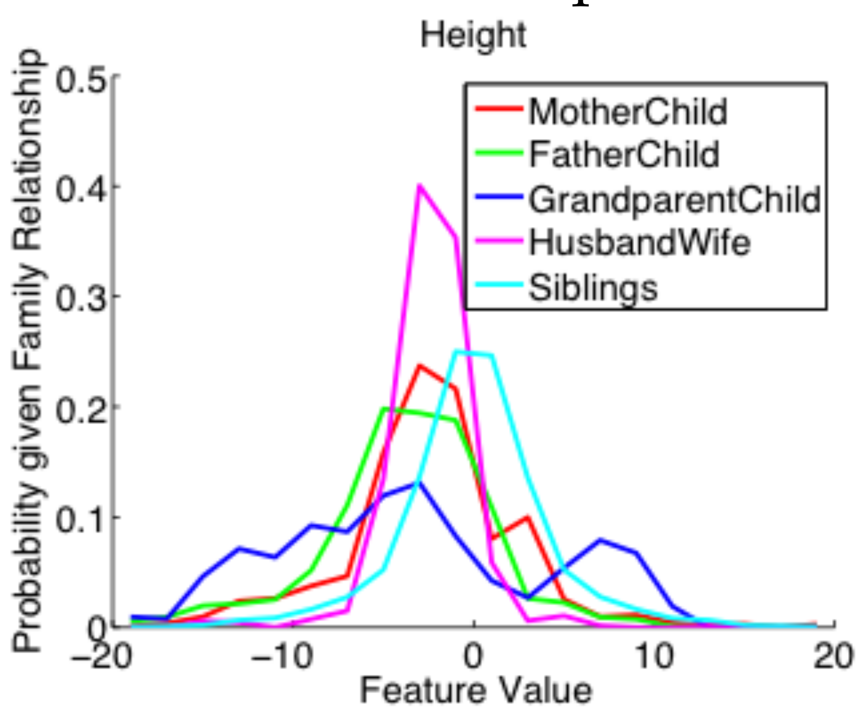
- Height difference
- Face size ratio
- Closeness
- Age difference (appearance based)
- Gender (appearance based)

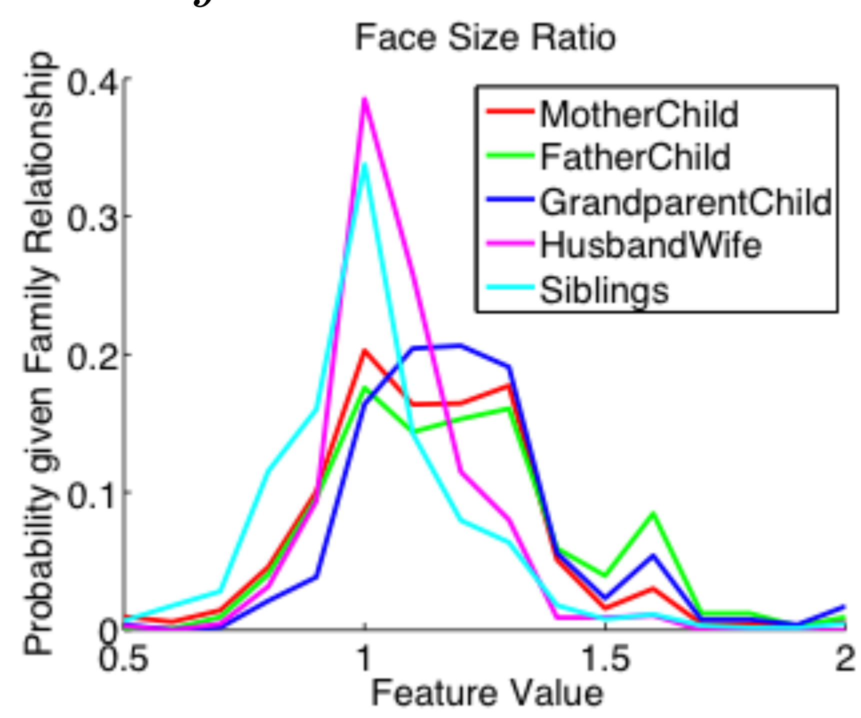
Representing Social Relationships

r_{ij}: social relationship between i^{th} and j^{th} person

mother-child	father-child	grandparent-child	husband-wife	siblings
child-mother	child-father	child-grandparent	wife-husband	

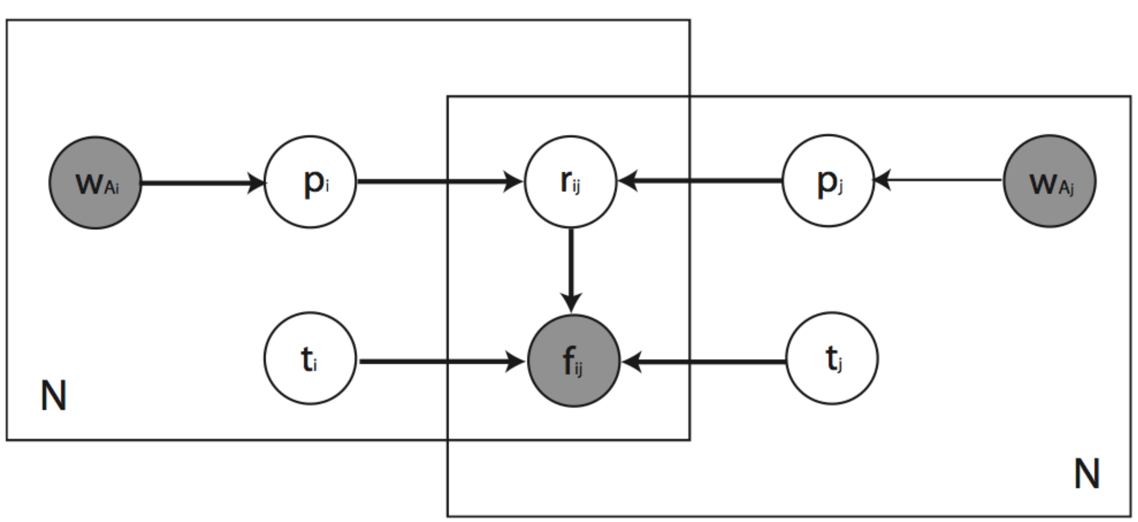
f_{ij}: social relationship 'features' between i^{th} and j^{th} face





Model

$$\sum_{A}\prod_{i=1}^{N} p(p_i \mid w_{A_i}) \prod_{i=1,j=1}^{N} p(f_{A_iA_j} \mid r_{ij}, t_i, t_j) p(r_{ij} \mid p_i, p_j) p(A)$$



p_i: the i^{th} person name

w_{Ai}: facial features associated with p_i

r_{ij}: social relationship between i^{th} and j^{th} person

t_i: age of the i^{th} person

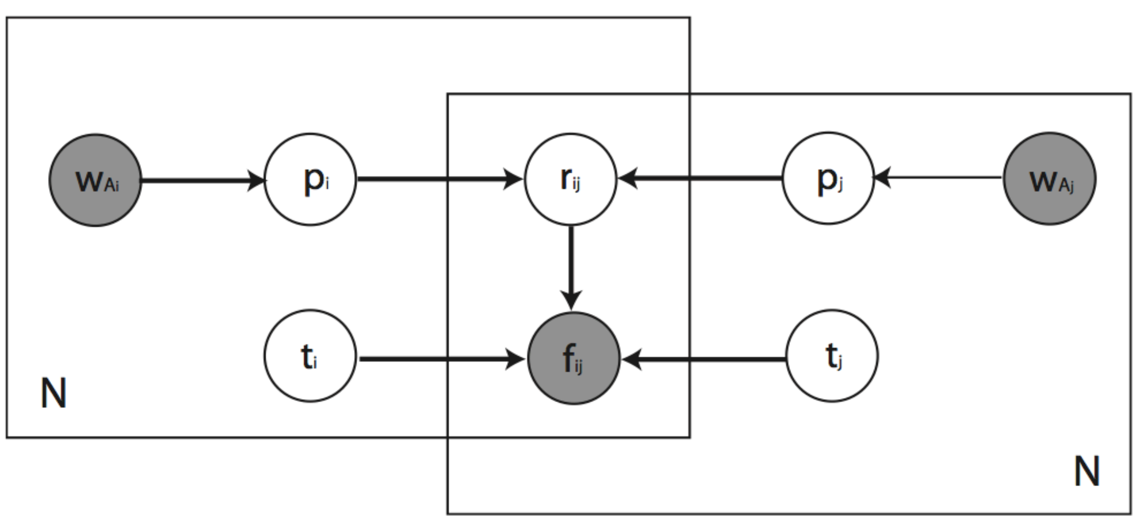
f_{ij}: social relationship 'features' between i^{th} and j^{th} face

Model

$$\boxed{\sum_{A}\prod_{i=1}^{N} p(p_i \mid w_{A_i})} \prod_{i=1,j=1}^{N} p(f_{A_iA_j} \mid r_{ij}, t_i, t_j)p(r_{ij} \mid p_i, p_j)p(A)$$

Appearance term represented with a discriminative model.

w_{Ai} denotes facial features associated with p_i



p_i: the i^{th} person name

w_{Ai}: facial features associated with p_i

r_{ij}: social relationship between i^{th} and j^{th} person

t_i: age of the i^{th} person

f_{ij}: social relationship 'features' between i^{th} and j^{th} face

Model

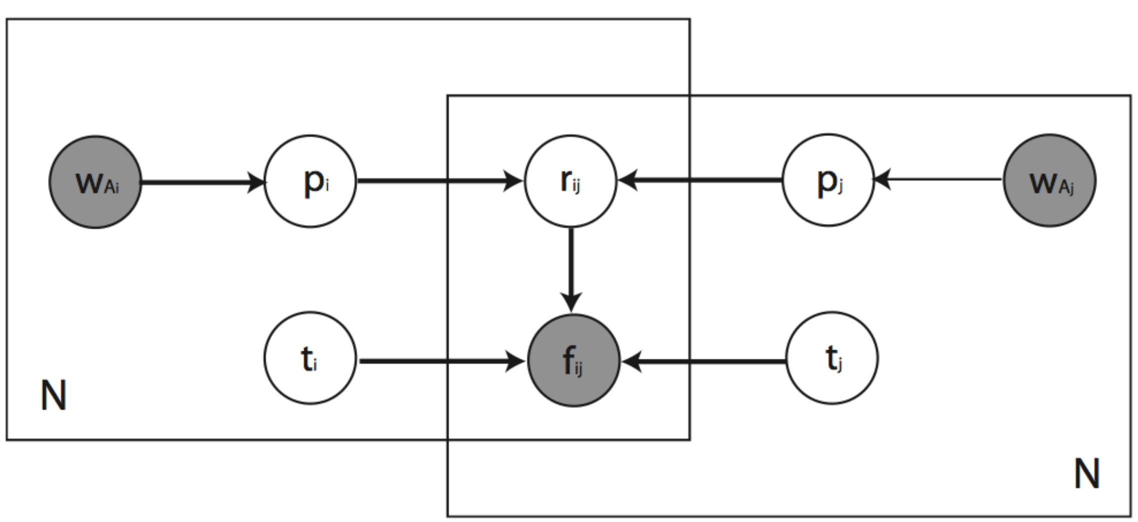
$$\sum_{A}\prod_{i=1}^{N} p(p_i \mid w_{A_i}) \boxed{\prod_{i=1,j=1}^{N} p(\underline{f_{A_iA_j}} \mid \underline{r_{ij}, t_i, t_j}) p(r_{ij} \mid p_i, p_j) p(\underline{A})}$$

Relationship term represented with a generative model.

f_{AiAj} denotes social relationship 'features' between faces A_i and A_j

r_{ij} denotes the discrete social relationship between i^{th} and j^{th} person

A is a hidden variable that relates names and faces



p_i: the i^{th} person name

w_{Ai}: facial features associated with p_i

r_{ij}: social relationship between i^{th} and j^{th} person

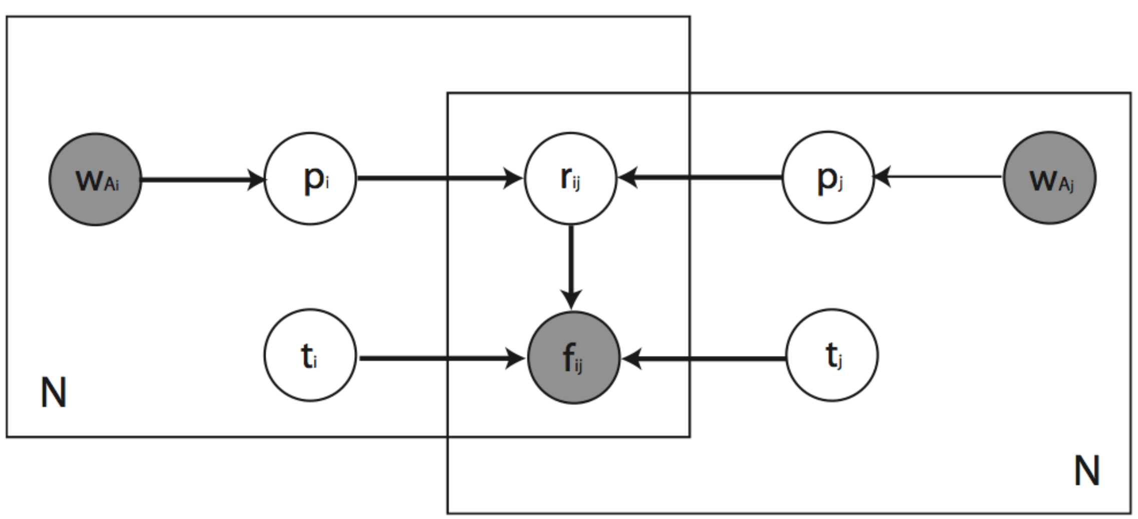
t_i: age of the i^{th} person

f_{ij}: social relationship 'features' between i^{th} and j^{th} face

Model

$$\sum_{A} \prod_{i=1}^{N} p(p_i \mid w_{A_i}) \boxed{\prod_{i=1,j=1}^{N} p(\underline{f_{A_iA_j}} \mid \underline{r_{ij}, t_i, t_j}) \boxed{p(r_{ij} \mid p_i, p_j)} \underline{p(A)}}$$

Since relationships are annotated $p(\ r_{ij} \mid p_i,\ \ p_j) = 1$



p_i: the i^{th} person name

w_{Ai}: facial features associated with p_i

r_{ij}: social relationship between i^{th} and j^{th} person

t_i: age of the i^{th} person

f_{ij}: social relationship 'features' between i^{th} and j^{th} face

Learning

Learning

Learn using EM

Learning

Learn using EM

Parameter: $$\widehat{\theta} = \mathrm{argmax}_{\theta} p(P, R, T \mid W, F; \theta)$$

Simplifications: System initialized with parameters produced by the baseline model (omits social relationships)

Learning

Learn using EM

Parameter: $\widehat{\theta} = \mathrm{argmax}_{\theta} p(P, R, T \mid W, F; \theta)$

Simplifications: System initialized with parameters produced by the baseline model (omits social relationships)

E Step:

$$p(A^* \mid P, R, T, W, F; \theta^{\mathrm{old}}) = \frac{p(P, R, T, W, F \mid A^*; \theta^{\mathrm{old}}) p(A^*; \theta^{\mathrm{old}})}{\sum_A p(P, R, T, W, F \mid A; \theta^{\mathrm{old}}) p(A; \theta^{\mathrm{old}})}$$

Simplifications: Prior distribution of A treated as uniform distribution. Only assign one p_i to a w_j when $\mathrm{p}(p_i \mid w_j)$ is bigger than a threshold.

Learning

Learn using EM

Parameter: $$\widehat{\theta} = \mathrm{argmax}_{\theta} p(P, R, T \mid W, F; \theta)$$

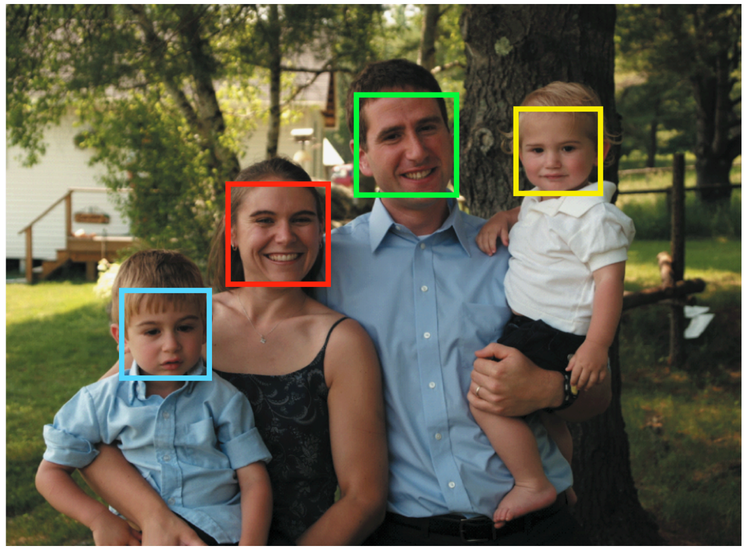
Simplifications: System initialized with parameters produced by the baseline model (omits social relationships)

E Step: $$p(A^* \mid P, R, T, W, F; \theta^{\mathrm{old}}) = \frac{p(P, R, T, W, F \mid A^*; \theta^{\mathrm{old}}) p(A^*; \theta^{\mathrm{old}})}{\sum_A p(P, R, T, W, F \mid A; \theta^{\mathrm{old}}) p(A; \theta^{\mathrm{old}})}$$

Simplifications: Prior distribution of A treated as uniform distribution. Only assign one p_i to a w_j when $\mathrm{p}(p_i \mid w_j)$ is bigger than a threshold.

M Step: Maximize by updating $p(p \mid w)$ and $p(f \mid r,\ t)$ separately.

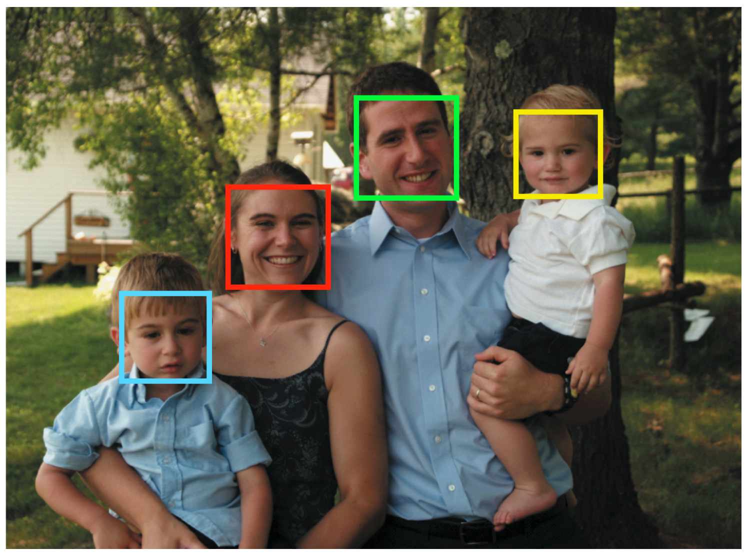
Inference



Input

- No name labels
- Extract facial features (W) and relationship features (F)

Inference



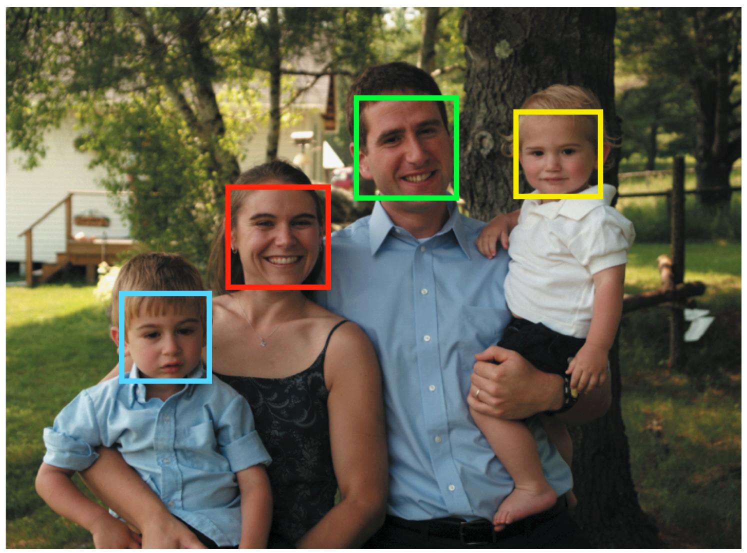
Input

$$A^{\star} = \mathrm{argmax}_A p(A \mid P, R, W, F, T)$$

Inference

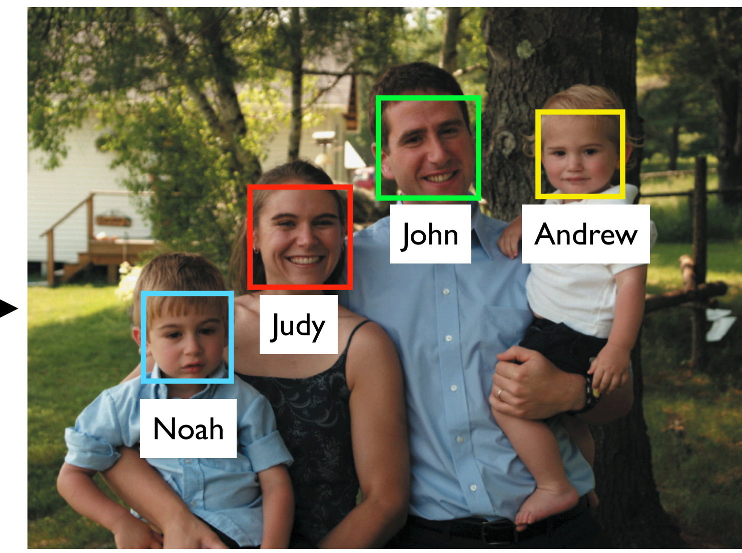
- No name labels
- Extract facial features (W) and relationship features (F)

Inference



$$A^\star = \mathrm{argmax}_A p(A \mid P, R, W, F, T)$$

Inference



Input

- No name labels
- Extract facial features (W) and relationship features (F)

Output

- Tagged faces (P)

Experiment A

Recognizing people with social relationships

Data

Training	Test

Collection 1:
1,125 images
47 people
600 training examples

Training	Test

Collection 2:
1,123 images
34 people
600 training examples

Training	Test

Collection 3:
1,117 images
152 people
600 training examples

Procedure

Annotate subset of data → Train model → Infer name assignment for test image

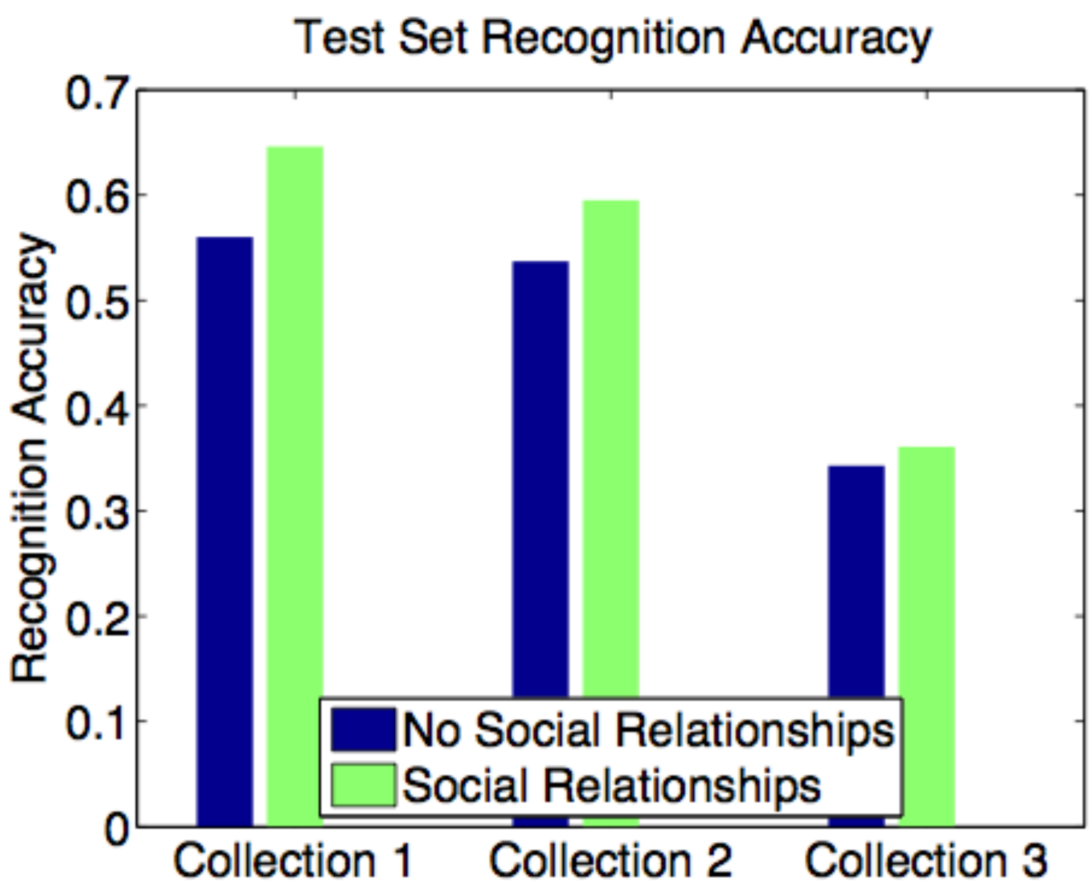
Results - Experiment A

Recognizing people with social relationships

Green - Correctly recognized with relationship modeling
Red - Incorrectly recognized with relationship modeling

Results - Experiment A

Recognizing people with social relationships



Average improvement:
5.4%

	without relationships	+height	+closeness	+size	+age	+gender	+all
Collection 1	0.560	0.621	0.628	0.637	0.635	0.630	**0.646**
Collection 2	0.537	0.563	0.560	0.583	0.573	0.584	**0.595**
Collection 3	0.343	0.361	0.359	**0.362**	0.362	0.362	0.361
Overall Mean	0.480	0.515	0.516	0.527	0.523	0.525	**0.534**

Experiment B

Recognizing social relationships in novel image sets

Data

Training	Test	Test
Collection 1: 1,125 images 47 people 600 training examples	Collection 2: 1,123 images 34 people	Public dataset [1]: 5,080 images 28,231 people

[1] A. Gallagher and T. Chen. Understanding Images of Groups of People. In Proc. CVPR, 2009.

Procedure

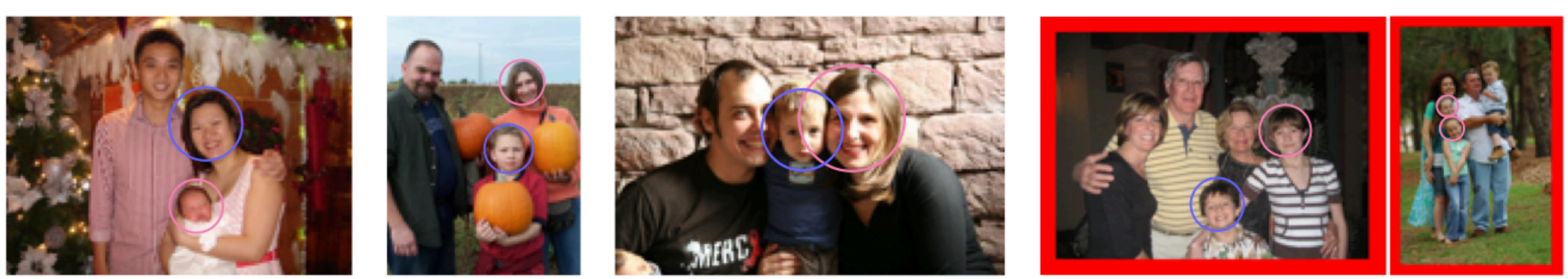
Train relationship model on Collection 1 → Classify social relationships on previously unseen image

Results - Experiment B

Recognizing social relationships in novel image sets

Husband-Wife

Siblings



Mother-Child

Results - Experiment B

Recognizing social relationships in novel image sets

Confusion Matrices

Test on Collection 2

	child–mother	mother–child	child–father	father–child	wife–husband	husband–wife	sibling
child–mother	.71	.01	.24			.01	.03
mother–child	.01	.71	.01	.24			.03
child–father	.39	.01	.32		.01	.23	.04
father–child	.01	.38		.33	.24	.01	.04
wife–husband	.07		.04	.01	.35	.05	.48
husband–wife		.07	.01	.03	.04	.40	.44
sibling	.07	.07	.02	.02	.05	.05	.73

Random assignment = 14.3%
Average Performance = 50.8%

Test on Public Collection

	child–mother	mother–child	wife–husband	husband–wife	sibling
child–mother	.52		.18	.09	.21
mother–child	.01	.65	.04	.25	.05
wife–husband	.15	.08	.52	.17	.08
husband–wife	.08	.16	.04	.53	.19
sibling	.06	.12	.15	.24	.42

Random assignment = 20%
Average Performance = 52.7%

Discussion

- Relatively large no. of training examples (50% of collection). What is the actual overhead of relationship labeling?
- Can we add more appearance based features?
  - Eg. Husband skin tone is darker than wife's *
- Performance of classifier in exceptional cases
  - Wife taller than husband
  - Same-sex couple
- Marginal improvement - 5.4%
  - They use Fisher subspace features (weak). Will the gain reduce if we include more attributes?
- Limited to family photos. Other applications?

* Manyam et. al. "Two faces are better than one", IJCB 2011

Thanks!