Microsoft Research Cambridge

Decision Tree Fields

Sebastian Nowozin
Microsoft Research Cambridge

Carsten Rother
Microsoft Research Cambridge

Shai Bagon
Weizmann Institute

Toby Sharp
Microsoft Research Cambridge

Bangpeng Yao
Stanford University

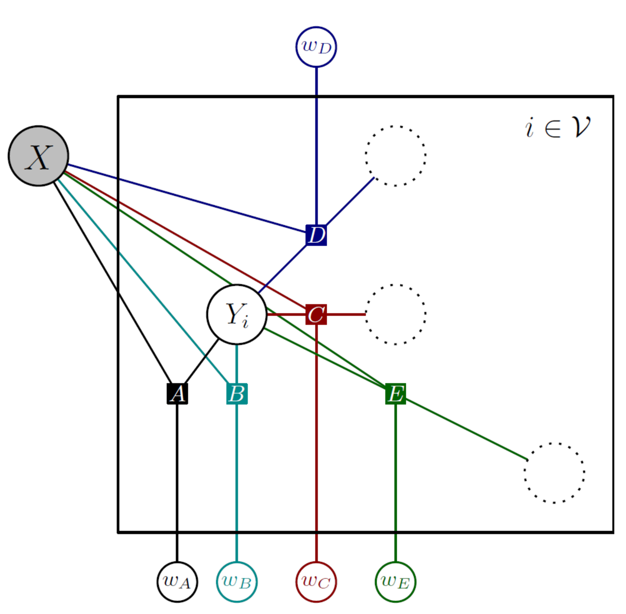
Pushmeet Kohli
Microsoft Research Cambridge

Overview

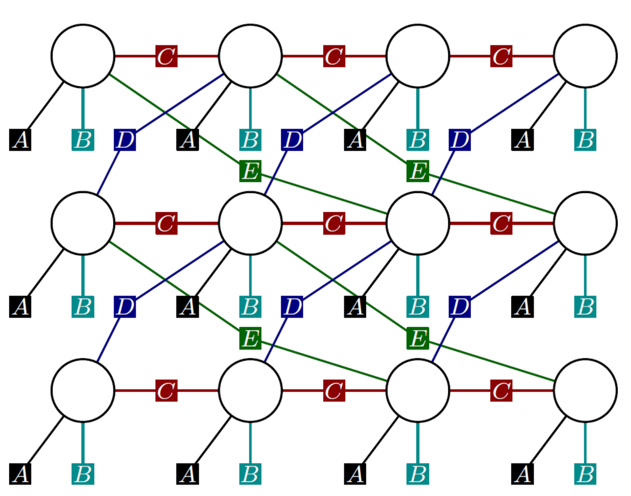
DTF = Efficiently learnable non-parametric CRFs for discrete image labelling tasks

- All factors (unary, pairwise, higher-order) are represented by decision trees
- Decision trees are *non-parametric*
- Efficient training of millions of parameters using pseudo-likelihood

Formally



Graphical Model: Factor types



Factor Graph

Energy $E(\boldsymbol{y},\boldsymbol{x},\boldsymbol{w}) = \sum_F E_{t_F}(y_F,\boldsymbol{x},w_{t_F})$

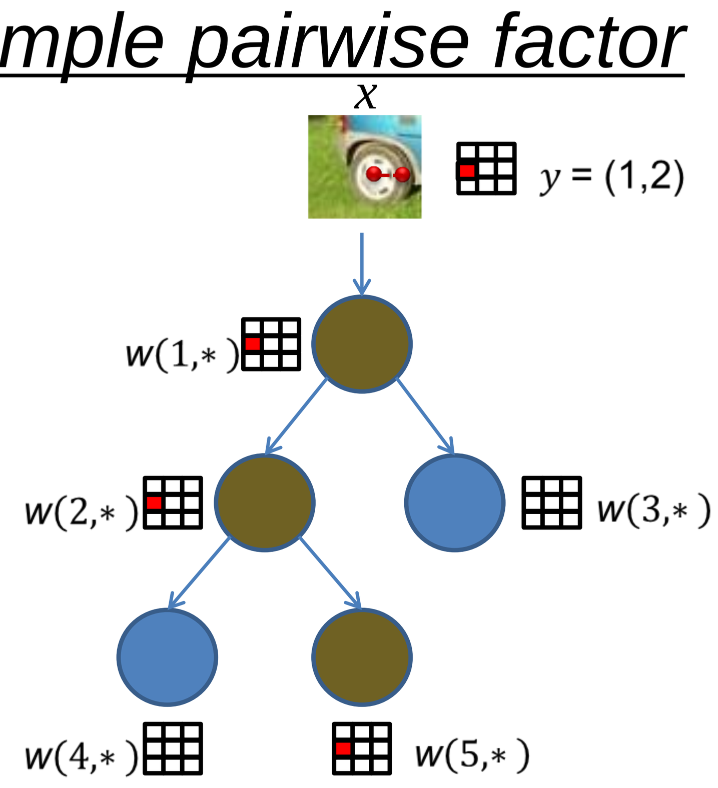
$$E_{t_F}(y_F,x,w_{t_F}) = \sum_{q \in Path(x_F)} w_{t_F}(q,y_F)$$

$$E_{t_F}(y_F,\boldsymbol{x},w_{t_F}) = \langle w_{t_F}, B_{t_F}(y_F,x_F)\rangle$$

$$B_{t_F}(y_F,x_F) =$$

Energy linear in w

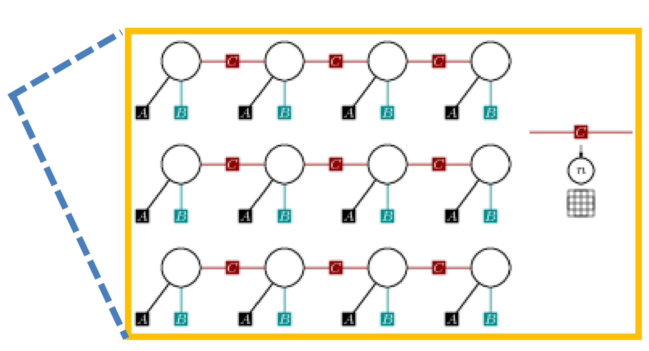
Example pairwise factor

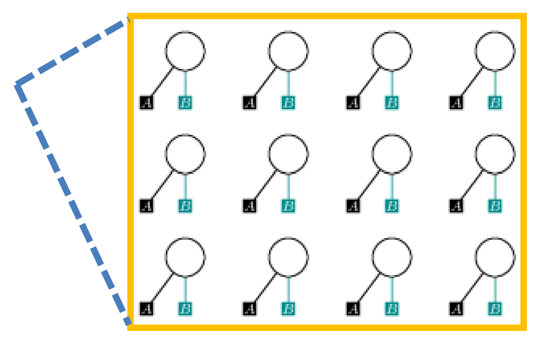


$E_{t_F}(y_F,\boldsymbol{x},\boldsymbol{w}_{t_F}) = w(1,(1,2)) + w(2,(1,2)) + w(5,(1,2))$

Special Cases

- Unary factors only = Decision Forest, with learned leaf node distributions
- Zero-depth trees (pairwise factors) = MRF
- Conditional (pairwise factors) = CRF





Algorithm - Overview

Training

1. Define connective structure (factor types)
2. Train all decision trees (split functions) separately
3. Jointly optimize all weights

Testing (2 options)

- "Unroll" factor graph:
  run: BP, TRW, QPBO, etc.
- Don't "unroll" factor graph:
  run Gibbs Sampling; Simulated Annealing

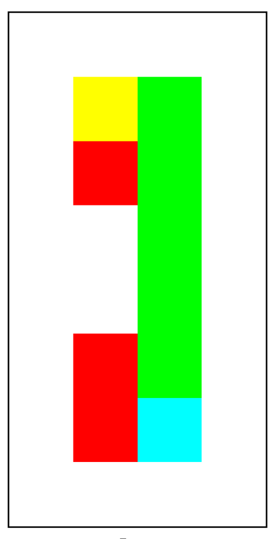
Training of weights "w"

- Maximum Pseudo-Likelihood training, convex optimization problem
- Converges in practice after 150-200 L-BFGS iterations
- Efficient even for large graphs (e.g. 12 connected, 1.47M weights, 22mins)
- Is parallel on the variable level
- Variable sub-sampling possible

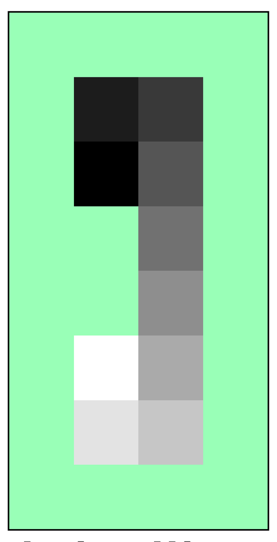
Code will be made available next month!

Results: Conditional Interactions - Snake Dataset

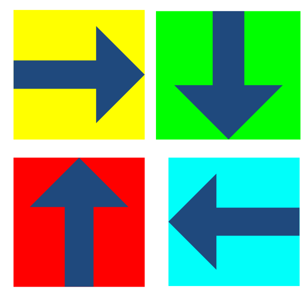
Training



Input image



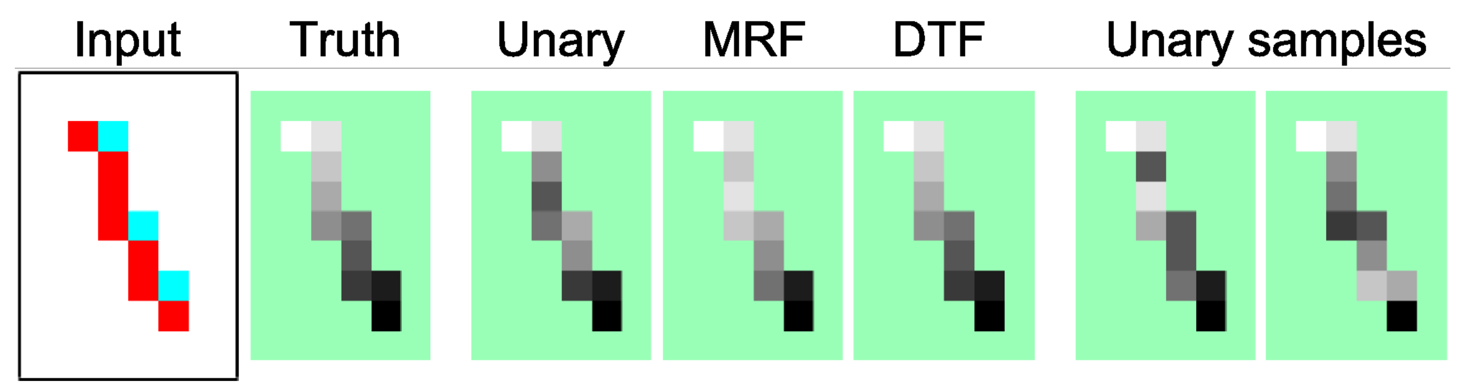
labelling



Colour encodes "direction"

200 randomly deforming snake images

Testing

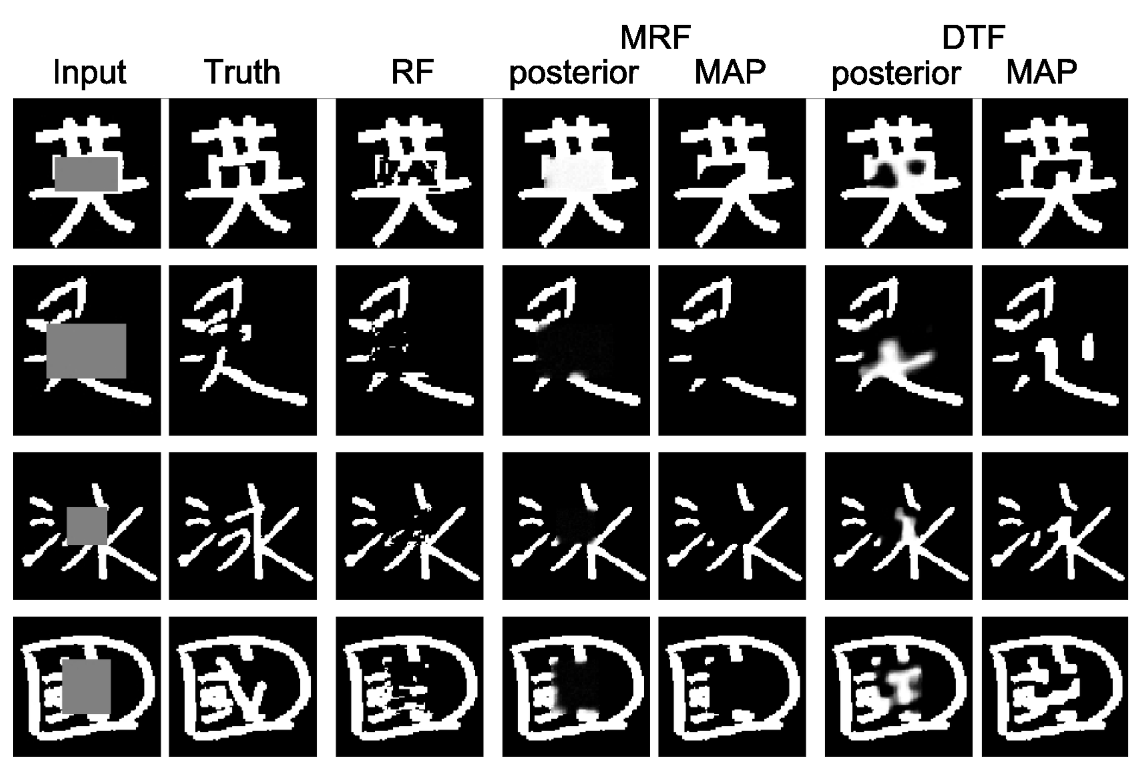


	RF	Unary	MRF	DTF
Avg. acc.	90.3%	90.9%	91.9%	**99.4%**
Tail acc.	100%	100%	100%	100%
Mid acc.	28%	28%	38%	**95%**

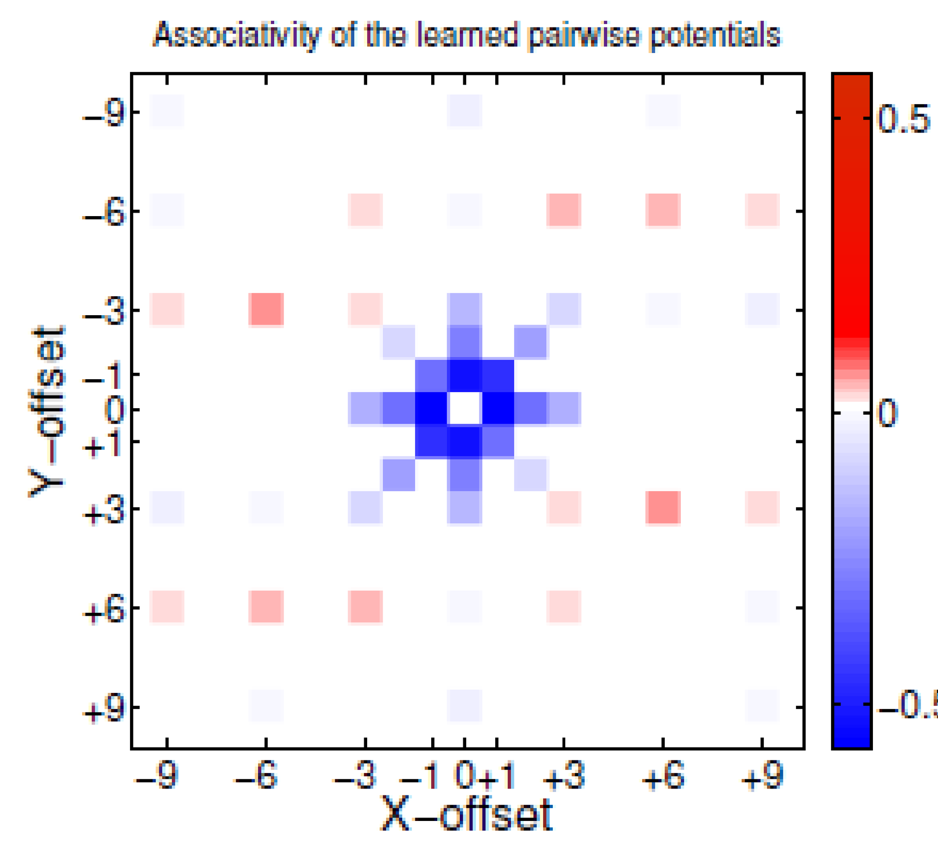
Conclusion: conditional pairwise terms are powerful

Results: Learning Calligraphy - Chinese Characters

300 Training images



100 Test images



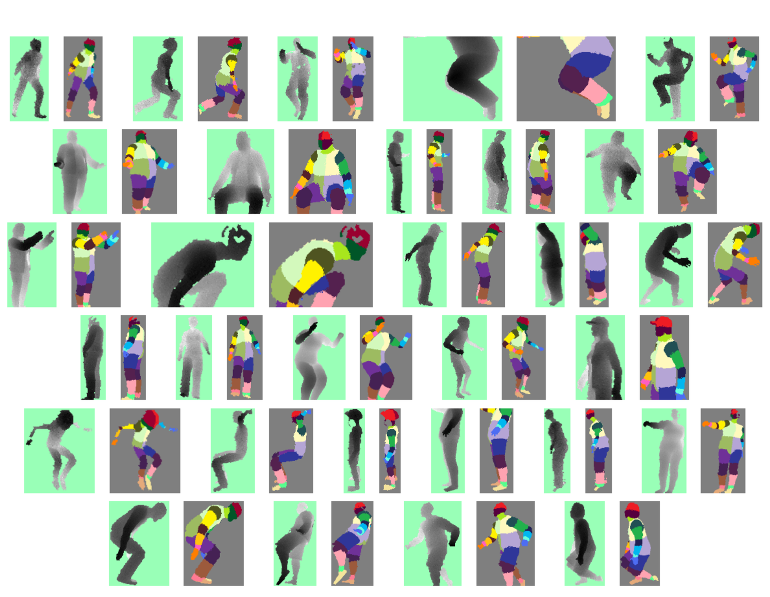
MRF weights

(blue attractive; red repulsive)

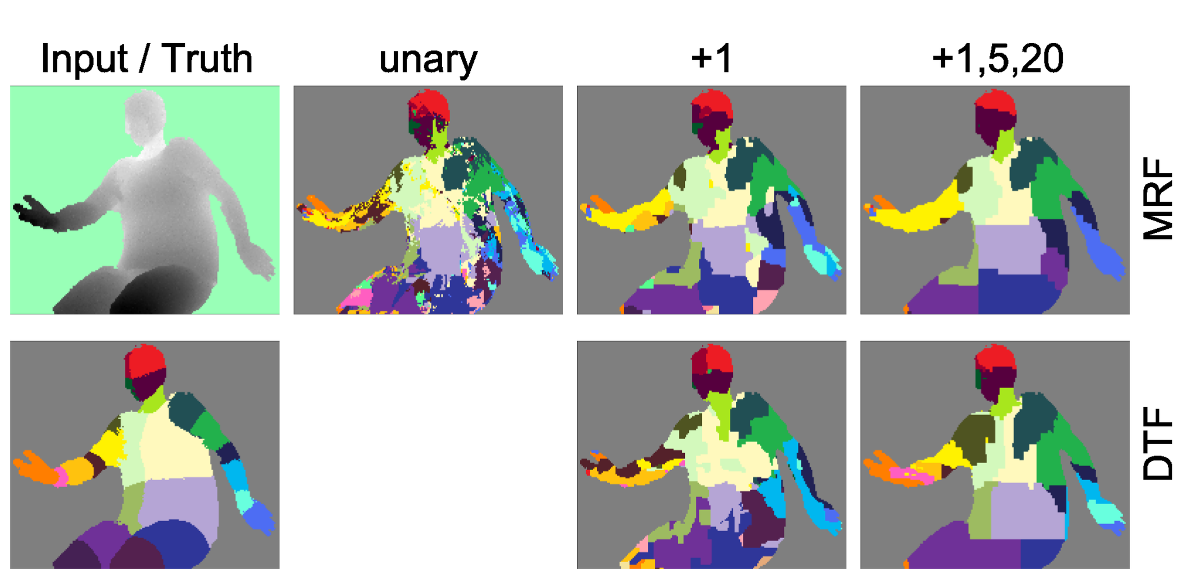
- **Densely connected pairwise DTF:**
  ~64 neighbours per variable, ~10k variables, ~300k factors, ~11k learned parameters
- **Test-time inference with simulated annealing (Gibbs chain)**
- **Hard energy minimization instances of this task are online:**
  http://www.nowozin.net/sebastian/papers/DTF_CIP_instances.zip

Results: Kinect-based bodypart detection

- **Body part recognition from depth images (Shotton et al., CVPR 2011)**
- **DTF: 4 unary factor types, 20 pairwise (+1,+5,+20)**
- **1500 training images, 150 test images**
- **Test-time inference with TRW (unrolled)**



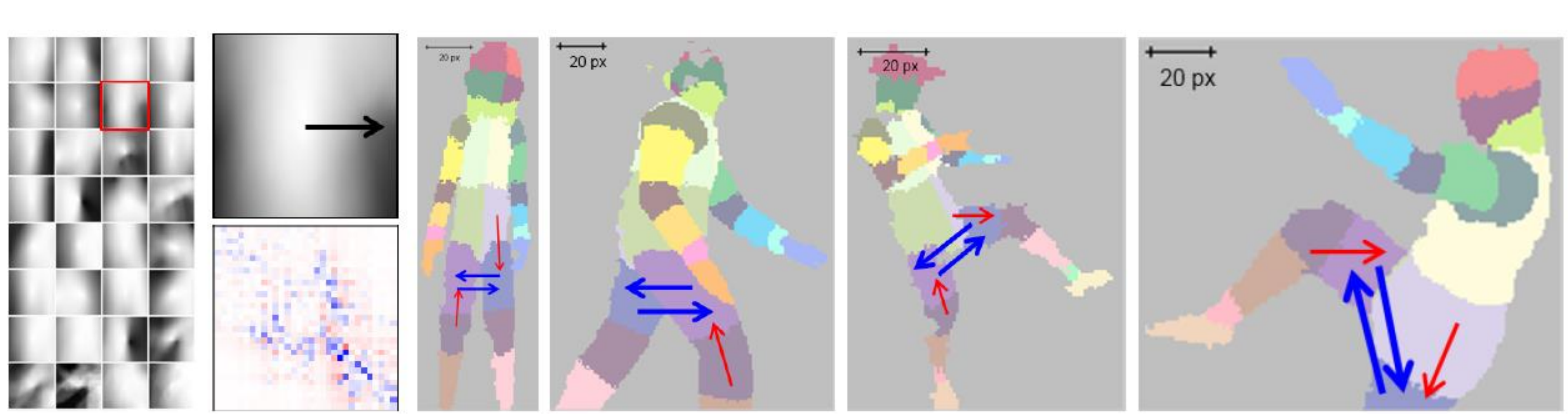
Example training images



Example test images

Model	Measure	[27]	unary	+1	+1,20	+1,5,20
MRF	avg-acc	14.8	21.36	21.96	23.64	24.05
30	runtime	1m	3m18	3m38	10m	10m
	weights	-	176k	178k	183k	187k
DTF	avg-acc	-	-	23.71	25.72	**27.35**
30	runtime	-	-	5m16	17m	22m
	weights	-	-	438k	951k	1.47M
MRF	avg-acc	34.4	36.15	37.82	38.00	39.30
1500	runtime	6h34	*	*	*	$(30h)*$
	weights	-	6.3M	6.2M	6.2M	6.3M
DTF	avg-acc	-	-	39.59	40.26	**41.42**
1500	runtime	-	-	*	*	$(40h)*$
	weights	-	-	6.8M	7.8M	8.8M

Test performance



Illustrating one learned horizontal interaction (20 pixels apart)