Introduction & Related Work

Motivation:
- Classical structured prediction methods cannot model global correlations efficiently.
- Newer methods can model global correlations efficiently but do not allow for known structure to be specified.

Prior Structured Output Prediction Approaches:

a) Independent
b) Structured

c) Deep Nets (a): Don’t model output variable correlation explicitly
- Explicitly model correlations between output variables
- Inference with high-order correlations scales poorly

Structured Deep Nets (b):
- Explicitly model correlations between output variables
- Inference with high-order correlations scales poorly

Structured Prediction Energy Networks (Bellanger & McCallum 2016) (c):
- Model global correlations among output variables
- Structural constraints not maintained during inference

Our approach (d):
- Model learned global structure and known local structure
- Structural constraints maintained during inference

Deep Nets (a):

Motivation:
- Classical structured prediction methods cannot model global correlations efficiently.
- Newer methods can model global correlations efficiently but do not allow for known structure to be specified.

Prior Structured Output Prediction Approaches:

a) Independent
b) Structured
c) Deep Nets (a): Don’t model output variable correlation explicitly
- Explicitly model correlations between output variables
- Inference with high-order correlations scales poorly

Structured Deep Nets (b):
- Explicitly model correlations between output variables
- Inference with high-order correlations scales poorly

Structured Prediction Energy Networks (Bellanger & McCallum 2016) (c):
- Model global correlations among output variables
- Structural constraints not maintained during inference

Our approach (d):
- Model learned global structure and known local structure
- Structural constraints maintained during inference

Deep Structured (b):
- Unary (a):
- Form Lagrangian w/ Lagrange multipliers λ

\[
\min_y \left( \frac{1}{2} \left( T\left(c, H(x, c, w), w\right) - \lambda^T y \right) + \max_{x \in X} \lambda^T H(x, c, w) \right)
\]

- Formulate max_x as ILP, relax to LP, form dual program \( H^D \)

\[
\min_{\mu} \left( \min_y \left( \frac{1}{2} \left( T\left(c, y, w\right) - \lambda^T y \right) + H^D(\mu, c, \lambda, w) \right) \right)
\]

- Solve w/ primal-dual algorithm from (Chambolle & Pock 2011)

Algorithm 1 Inference Procedure

Input: Learning rates \( \alpha \), \( \lambda \); number of iterations \( n \)
- Function \( f_j \) per output variable \( k \); simple sum of scores
- DeepStruct (b): \( f_j(x, c, w) = \sum_{r \in R} f_j(x, c, w) \)
- Function \( f_j \) per graph region \( r \); simple sum of scores
- Our Model (d): \( f_j(x, c, w) = T(c, H(x, c, w), w) \)
- Vector \( H \) contains one entry per assignment \( x \), per \( r \) in \( R \)
- Function \( T \) can be nonlinear gives global score based on local scores in \( H \)

Learning Objective: SSVM objective with loss-augmented inference

\[
\min_w \sum_{(c, e) \in D} \left( \max_{x \in X} \left( T(c, H(x, c, w), w) + L(x, \hat{x}) \right) - T(c, H(x, c, w), w) \right) + \frac{C}{2} \|w\|^2
\]