Optical Music Recognition via Image Scene Understanding

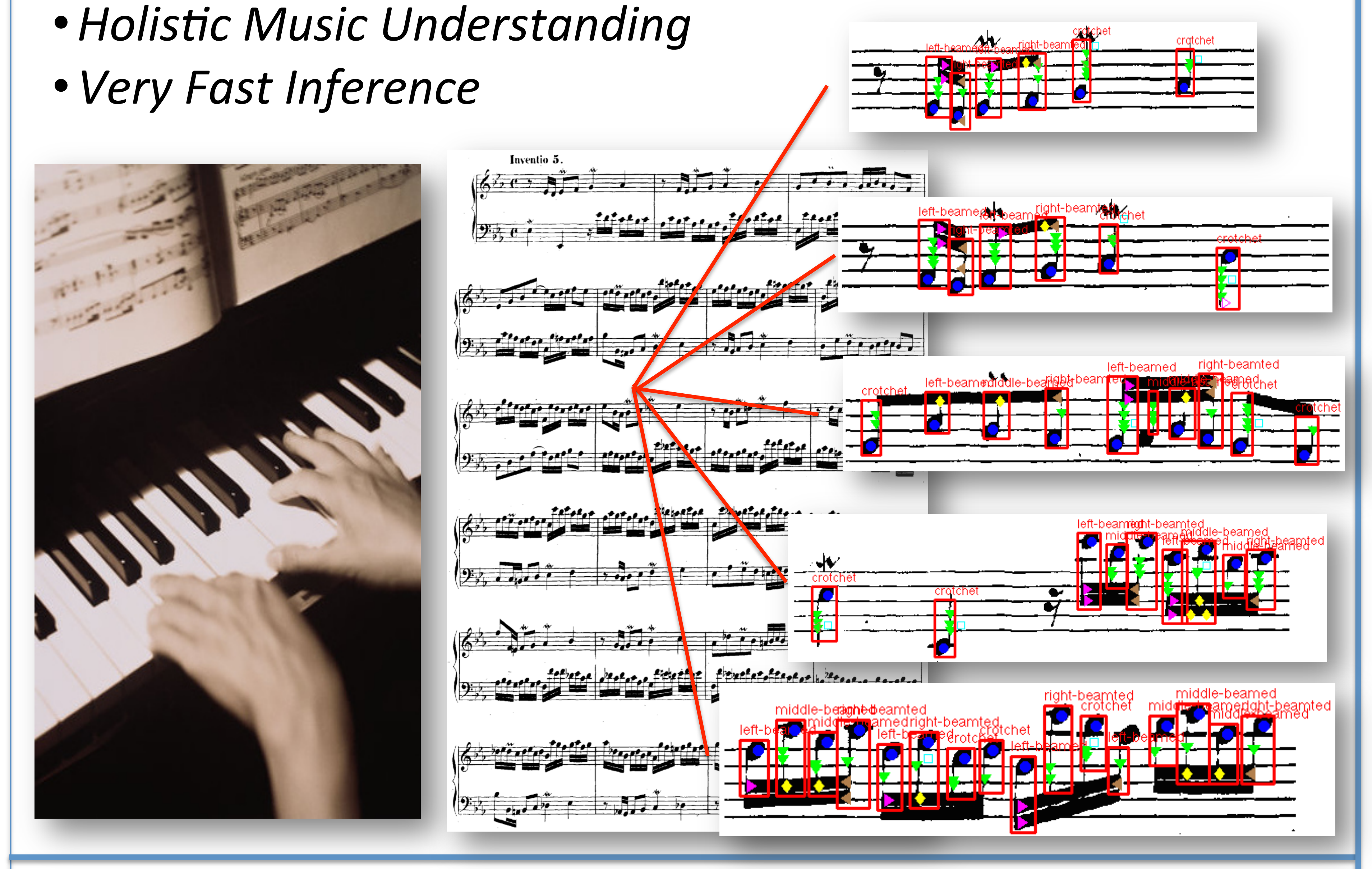
Liang Chen[1], Kun Duan[1], David Crandall[1]

[1]School of Informatics and Computing, Indiana University, Bloomington, IN

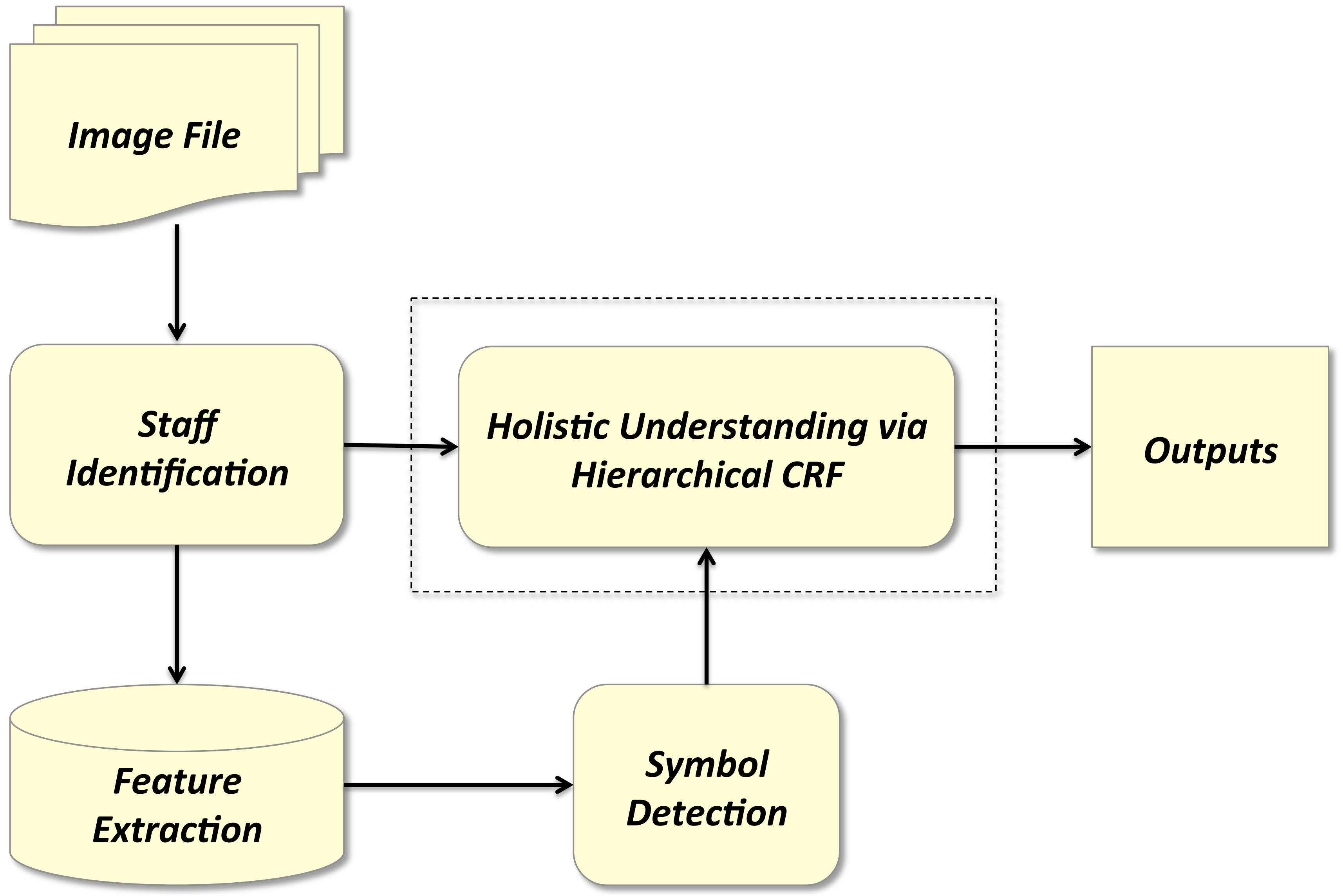
INDIANA UNIVERSITY

1. Overview

- *State-of-art Computer Vision Methods*
- *Holistic Music Understanding*
- *Very Fast Inference*

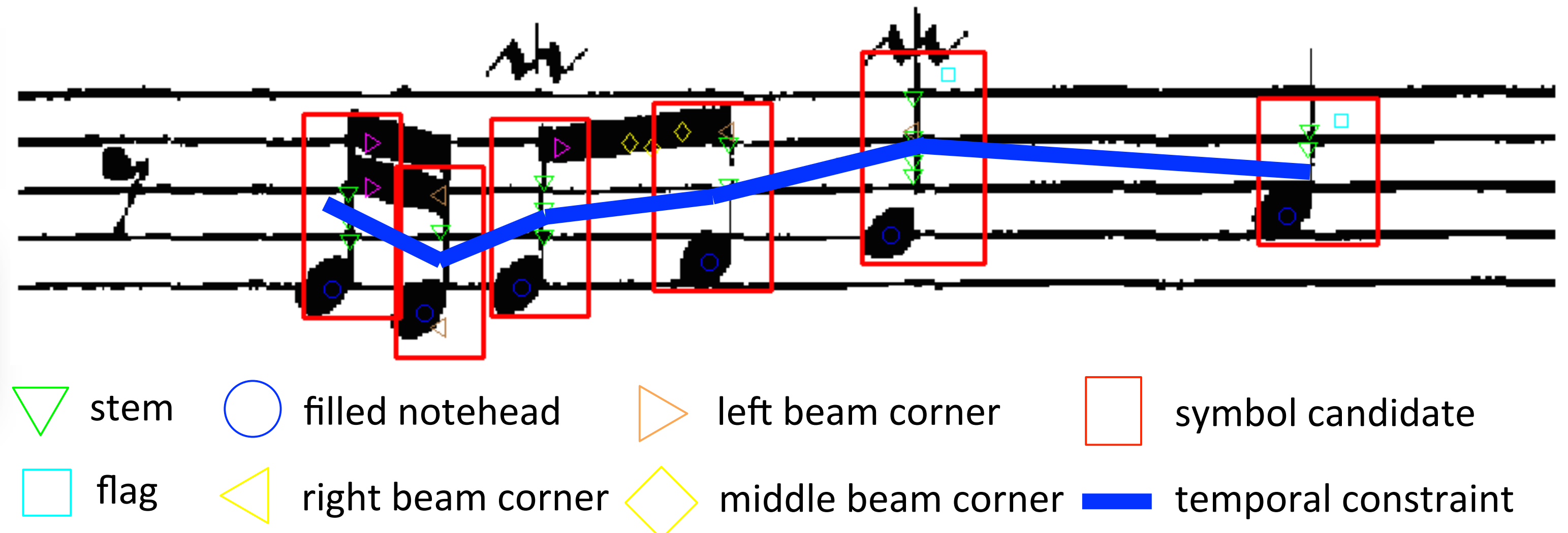


2. Optical Music Recognition Workflow



3. Modeling Music Sequence Using Hierarchical CRF

- *Measures are independent of each other*
- *A two-layer tree-structured CRF model*
- *Part detections as binary variables*
- *Find semantic symbols via mean-shift grouping*



part detection score

part-symbol coherence score

$$E(\{b_s\},\{y_t\}|X) = \sum_s \lambda(b_s|X) + \sum_t \mu(y_t|X) + \sum_{s,t} \phi(b_s^{(t)}, y_t|X) + \sum_{i,j} \psi(y_i, y_j|X)$$

symbol confidence

symbol-symbol coherence score

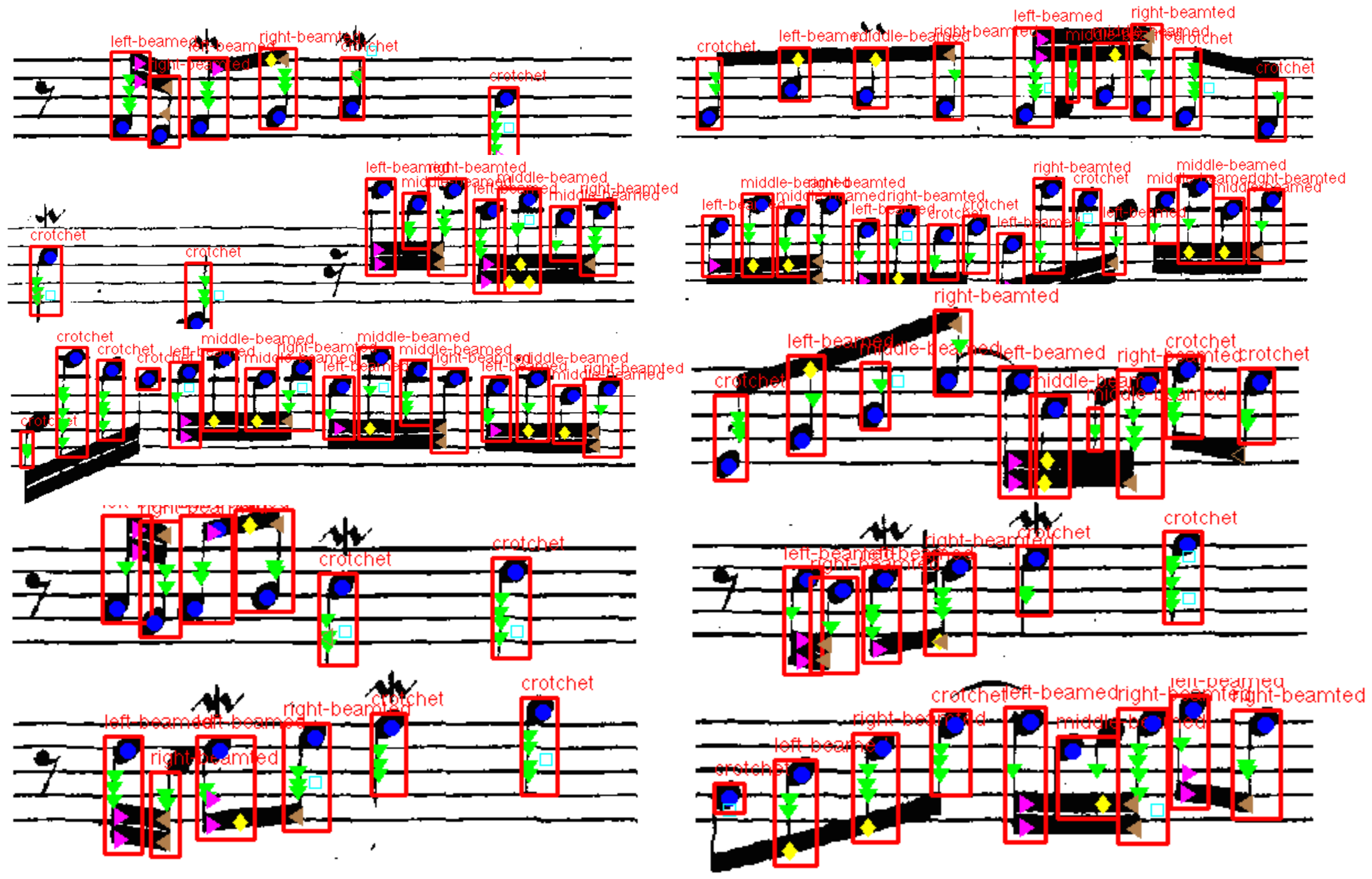
4. Structural SVM Learning

- *CRF parameters stacked into a single vector*
- *Part Loss defined using Intersection-over-Union*
- *Symbol loss defined using classification error*
- *Dual-coordinate descent algorithm for optimization*

$$\min_{\mathbf{w}} \|\mathbf{w}\|^2 + C\sum_{n=1}^{N} \left(\max_{\mathbf{y},\mathbf{b}} \left(\overbrace{\alpha_0\Delta_0(\mathbf{y}^{(n)},\mathbf{y}) + \alpha_1\Delta_1(\mathbf{b}^{(n)},\mathbf{b})}^{\text{weighted combined loss function}} + \mathbf{w}^T\Phi(\mathbf{X}^{(n)},\mathbf{y},\mathbf{b}) \right) - \mathbf{w}^T\Phi(\mathbf{X}^{(n)},\mathbf{y}^{(n)},\mathbf{b}^{(n)}) \right)$$

5. Results

- *Image Features:* Histogram-of-Gradients (HoG)
- *Dataset*: Johann Bach's Inventions
- *Annotation:* Pixel level labels for parts



6. Conclusions

- *Holistic music understanding gives better results than individual detections*
- *Tree-structured image scene model allows efficient inference*
- *Combined loss function captures loss with different characteristics*
- *A new benchmark dataset for optical music recognition problems*

7. Future work

- *Hand-written music recognition*
- *Musical document retrieval*
- *Automatic music generation*

COMPUTER VISION LAB

Music