Music Score Image

Guitar Pro (GPX)

converter

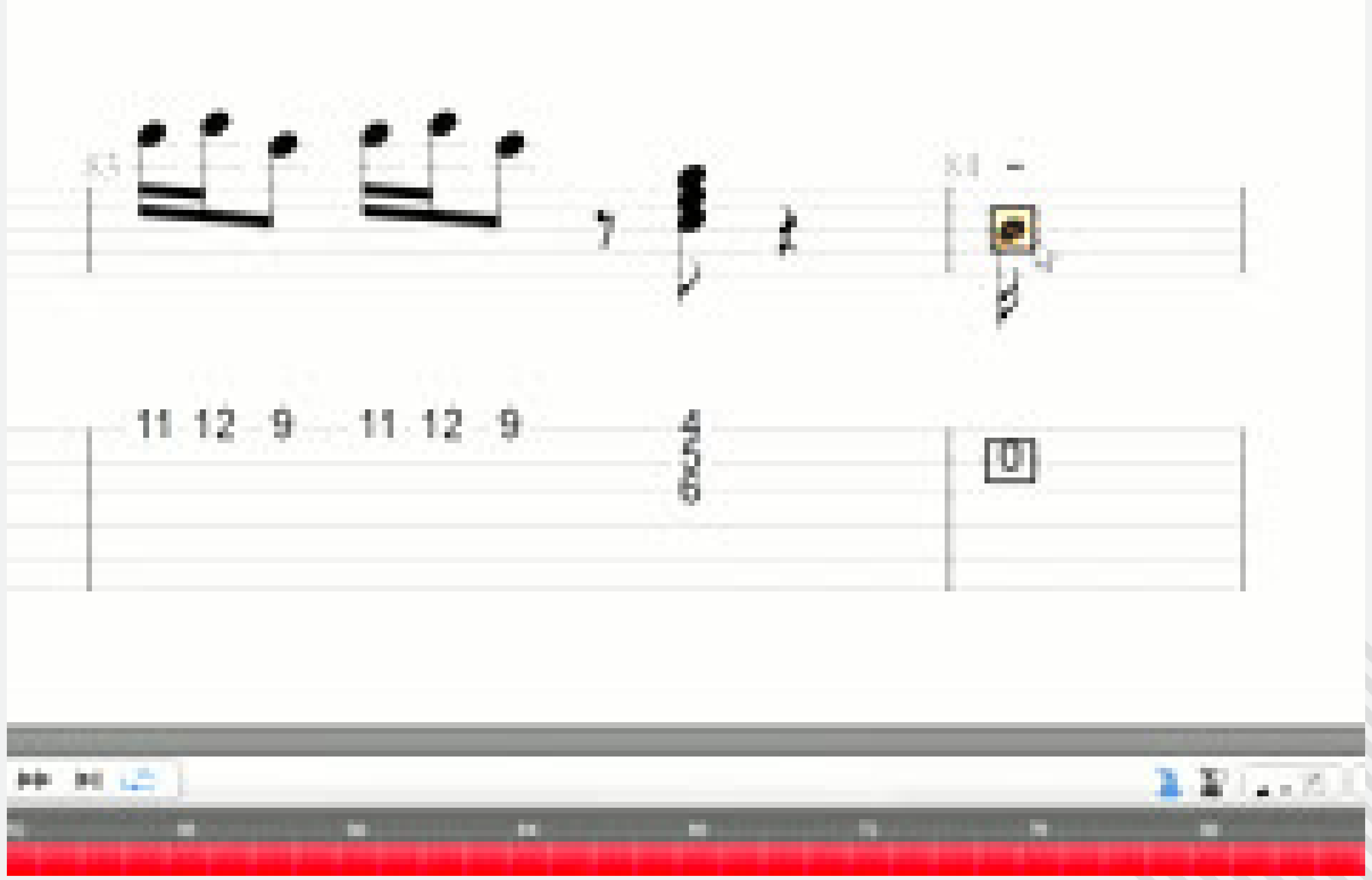
2018038015 홍석균
2018038037 이정렬
2019038054 김경민

COURSE: ARTIFICIAL INTELLIGENCE

Chungbuk National University

Contents

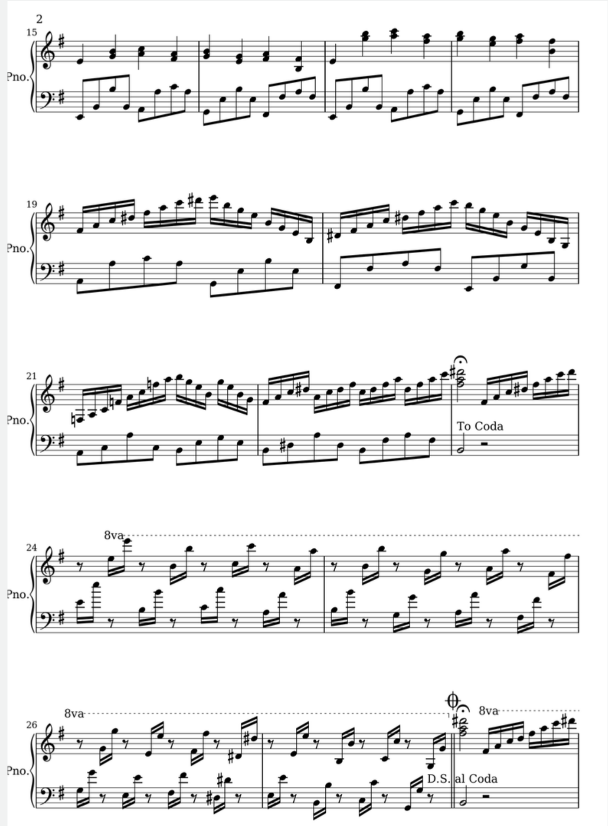
MOTIVATION



Identifying music symbols and **entering** them into a guitar pro program

➞ Music Scores Image Guitar Pro converter

RESEARCH OBJECT



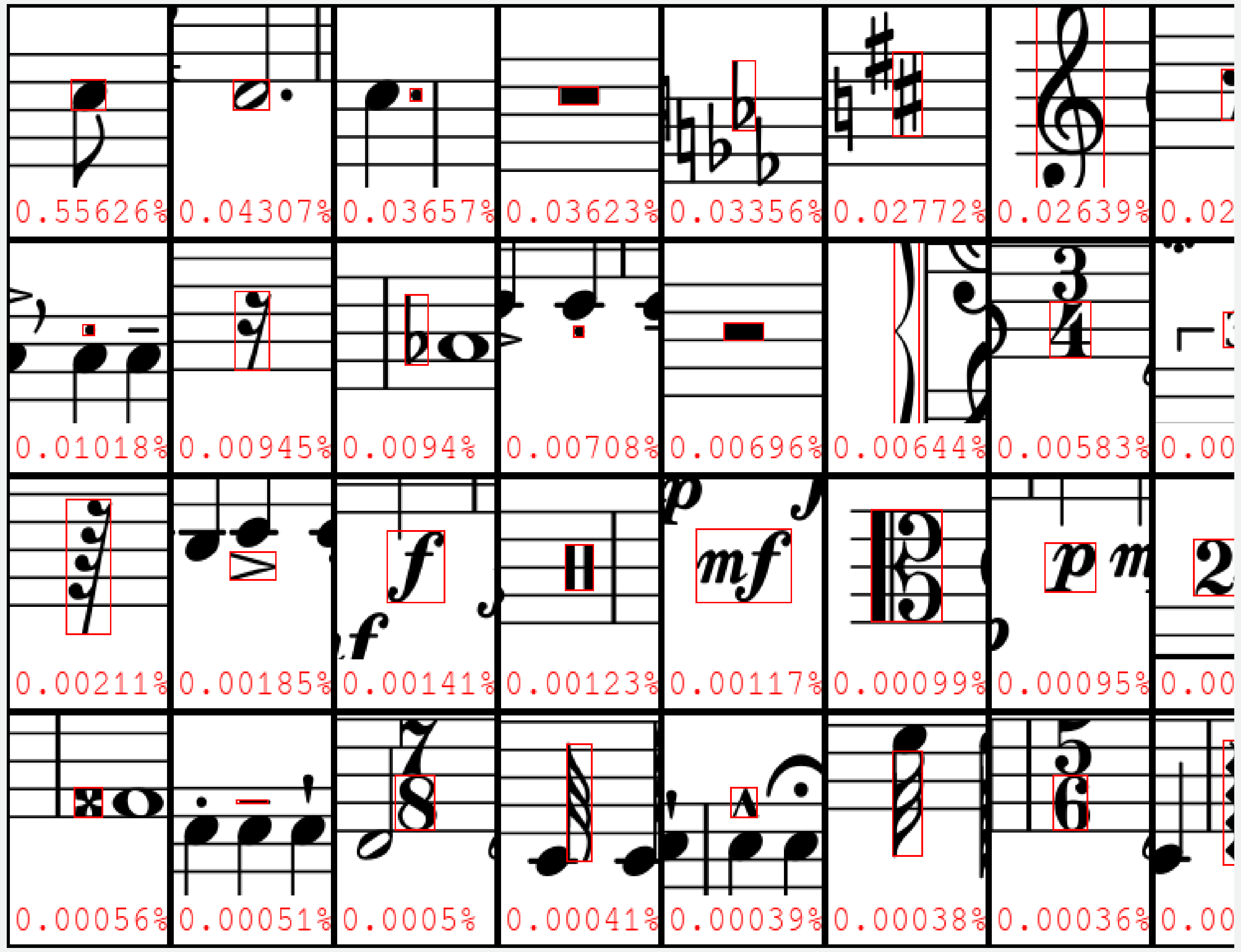
Music Scores Image

Object Detection

Create Guitar Pro file (.gpx)

MATERIAL (DATASET)

What is DeepScoresV2?



Music score image dataset Symbols of various shapes and sizes in Coco data format
(clef, key signature, note, rest, beat, scale) included

Ref. https://paperswithcode.com/paper/deepscores-a-dataset-for-segmentation

MATERIAL (DATASET)

Why DeepScoresV2?

Dataset	Classes	Images	Object Inst.	Avg. Inst. per Image
PASCAL VOC [19]	20	21,503	62,199	2.89
COCO 2014 [20]	80	123,287	886,266	7.19
ImageNet [21]	200	349,379	478,806	1.37
DOTA [22]	15	2,806	188,282	67.10
MUSCIMA++V2 [23]	163	140	102,914	735
DeepScoresV2	136	255,385	151M	592
↪ dense	136	1,714	1.1M	660

Select a dataset that fits the topic of recognizing music symbols through music score images
DeepScoresV2 dataset has a larger number of images and instances than MUSCIMA++V2

Ref. The DeepScoresV2 Dataset and Benchmark for Music Object Detection

MATERIAL (DATASET)

Structure of the dataset

▼ images [1]
  ▼ 0 {5}
    id : 410
    filename : lg-5230237-aug-emmentaler--page-2.png
    width : 2707
    height : 3828
    ▼ ann_ids [547]
      0 : 41408
      1 : 41409
      2 : 41410
      3 : 41411
      4 : 41412
      5 : 41413
      6 : 41414
      7 : 41415
      8 : 41416
      9 : 41417
      10 : 41418
      11 : 41419

▼ 41429 {6}
  ▼ a_bbox [4]
    0 : 1942.07250684
    1 : 746.48721708
    2 : 1945.02473397
    3 : 822.14158884
  ▼ o_bbox [8]
    0 : 1944.072509765625
    1 : 821.4872436523438
    2 : 1944.072509765625
    3 : 751.4872436523438
    4 : 1942.072509765625
    5 : 751.4872436523438
    6 : 1942.072509765625
    7 : 821.4872436523438
  ▼ cat_id [2]
    0 : 42
    1 : 161
  area : 213
  img_id : 410

categories {208}
  ▼ 1 {3}
    name : rest8th
    annotation_set : deepscores
    color : 1
  ▼ 2 {3}
    name : stem
    annotation_set : deepscores
    color : 2
  ▼ 3 {3}
    name : dynamicP
    annotation_set : deepscores
    color : 7

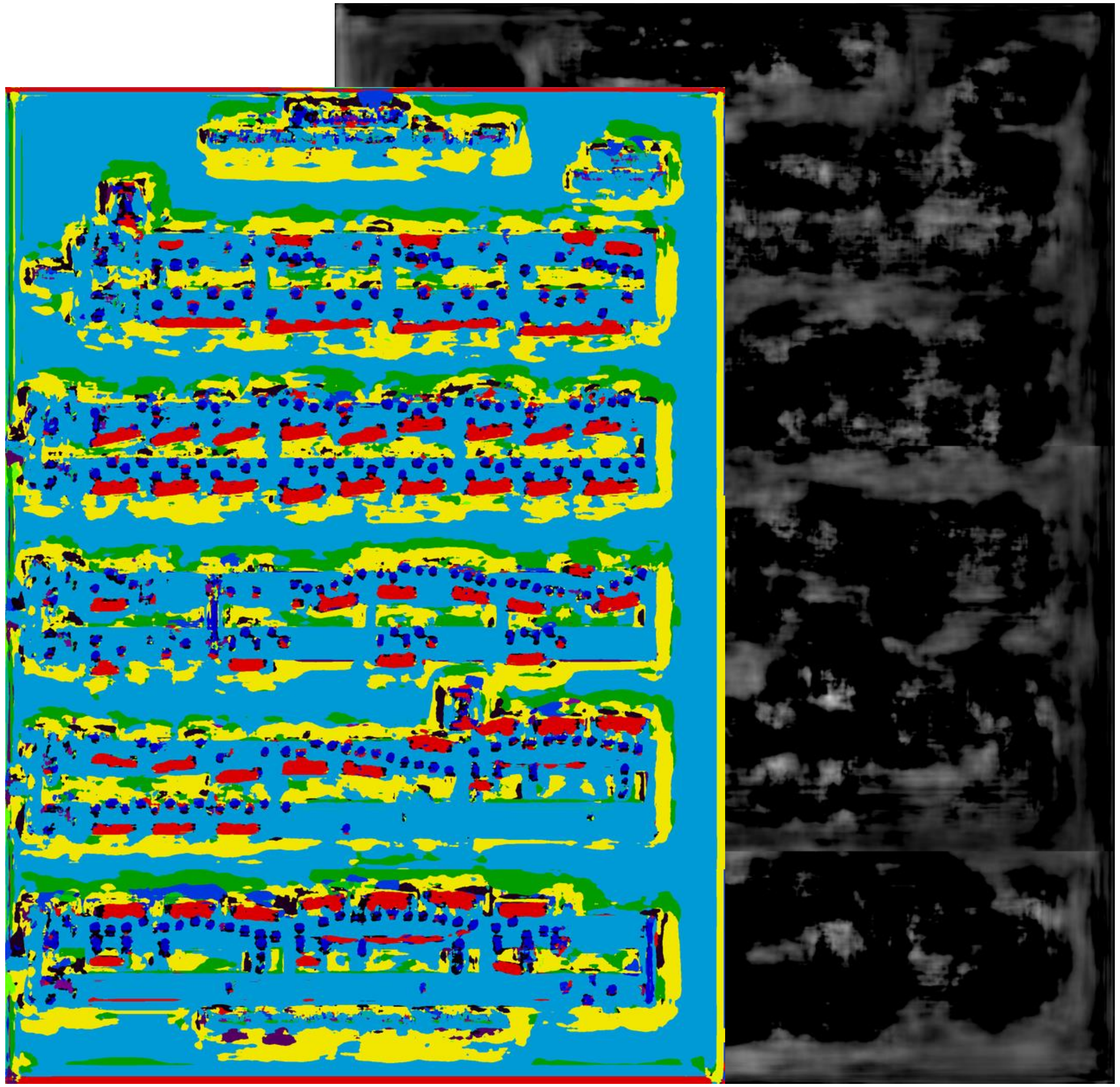
Deepscores V2 is in Coco data format and includes information about image, annotation (object), and category.
Consists of images, '**train.json**' for learning and '**test.json**' for testing.

MODEL

DEEP WATERSHED DETECTION

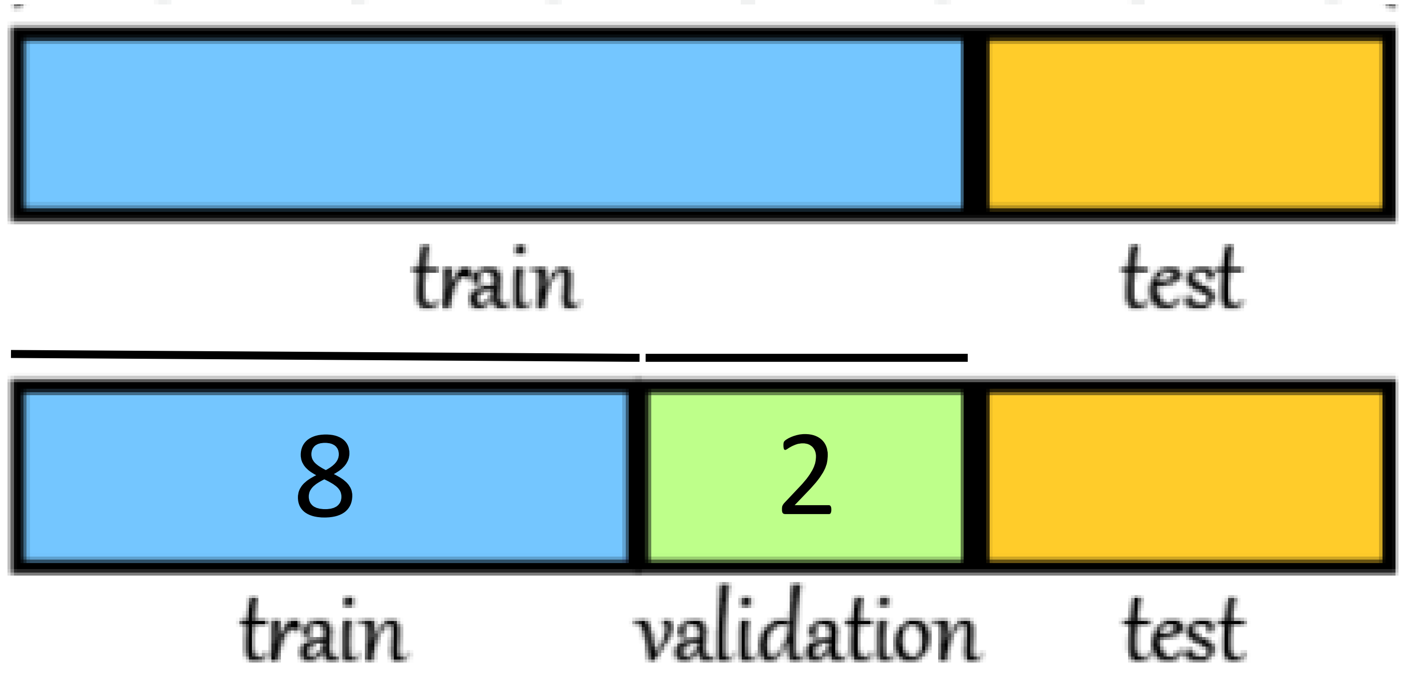
Watershed algorithm, one of the image segmentation algorithms, combined with deep learning network

Deep learning network identifies object boundaries in images
➡ Apply watershed transformation to perform segmentation



Ref. https://github.com/tuggeluk/DeepWatershedDetection

TRAINING

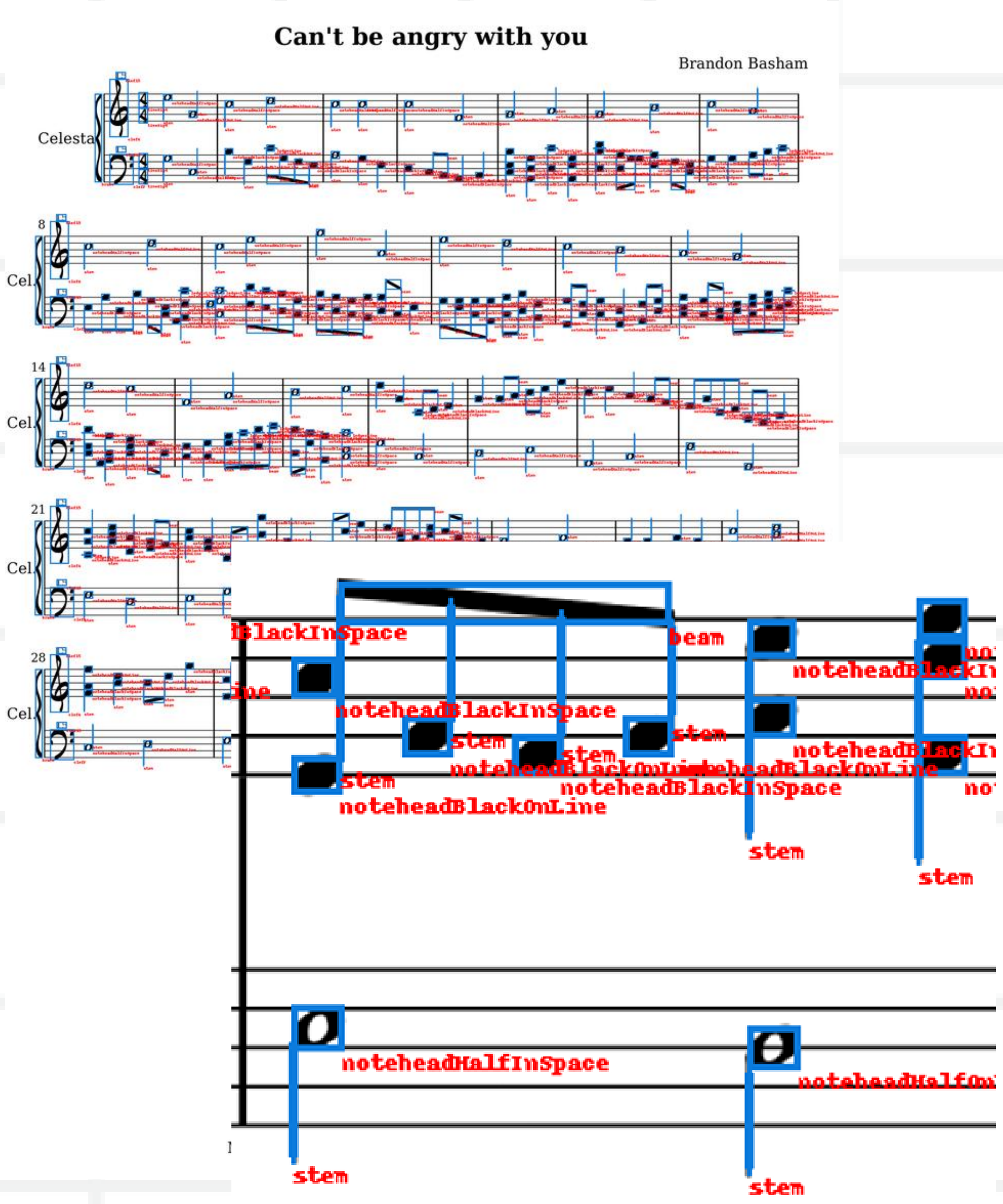


학습 횟수	iteration1	iteration2	iteration3	iteration4	iteration combine
1	10	10	10	20	30
2	10	10	10	20	30
3	50	50	50	40	60
4	70	70	70	60	100

5 steps per learning
batch size 1

RESULT

prediction



ground truth

RESULT

Nr	class	No. Occurences	mAP	AP at 0.5
0	ledgerLine	8857	0	0
1	clefG	820	2.51965e-07	2.51965e-06
2	clefF	534	0.000706414	0.00373715
3	noteheadBlackOnLine	13845	0.000264698	0.00211199
4	noteheadBlackInSpace	13635	0.0001671	0.00142662
5	noteheadHalfOnLine	957	0.000318644	0.00214455
6	noteheadHalfInSpace	1045	0.000158156	0.000834523
7	noteheadWholeOnLine	389	0	0
8	noteheadWholeInSpace	391	0	0
9	augmentationDot	1859	0	0
			0	0
			5.79406e-05	0.000320882
			0	0
			1.12023e-07	1.12023e-06
			0	0
			0	0
20	restWhole	753	7.22017e-05	0.000657758
21	restQuarter	736	0.000106242	0.000954252
22	rest8th	818	0.000129629	0.00087739
23	rest32nd	30	0	0
24	dynamicP	289	0.000110467	0.000941411
25	dynamicM	265	0	0
26	dynamicF	622	8.471e-05	0.000568641
27	dynamicS	37	0	0
28	stringsDownBow	29	0	0
29	slur	956	0	0
30	beam	7716	2.21617e-05	4.63631e-06
31	tie	1058	0	0
32	dynamicDiminuendoHairpin	53	0	0

$$\text{IoU} = \frac{\text{Area of Overlap}}{\text{Area of Union}}$$

mAP : Mean Average Precision

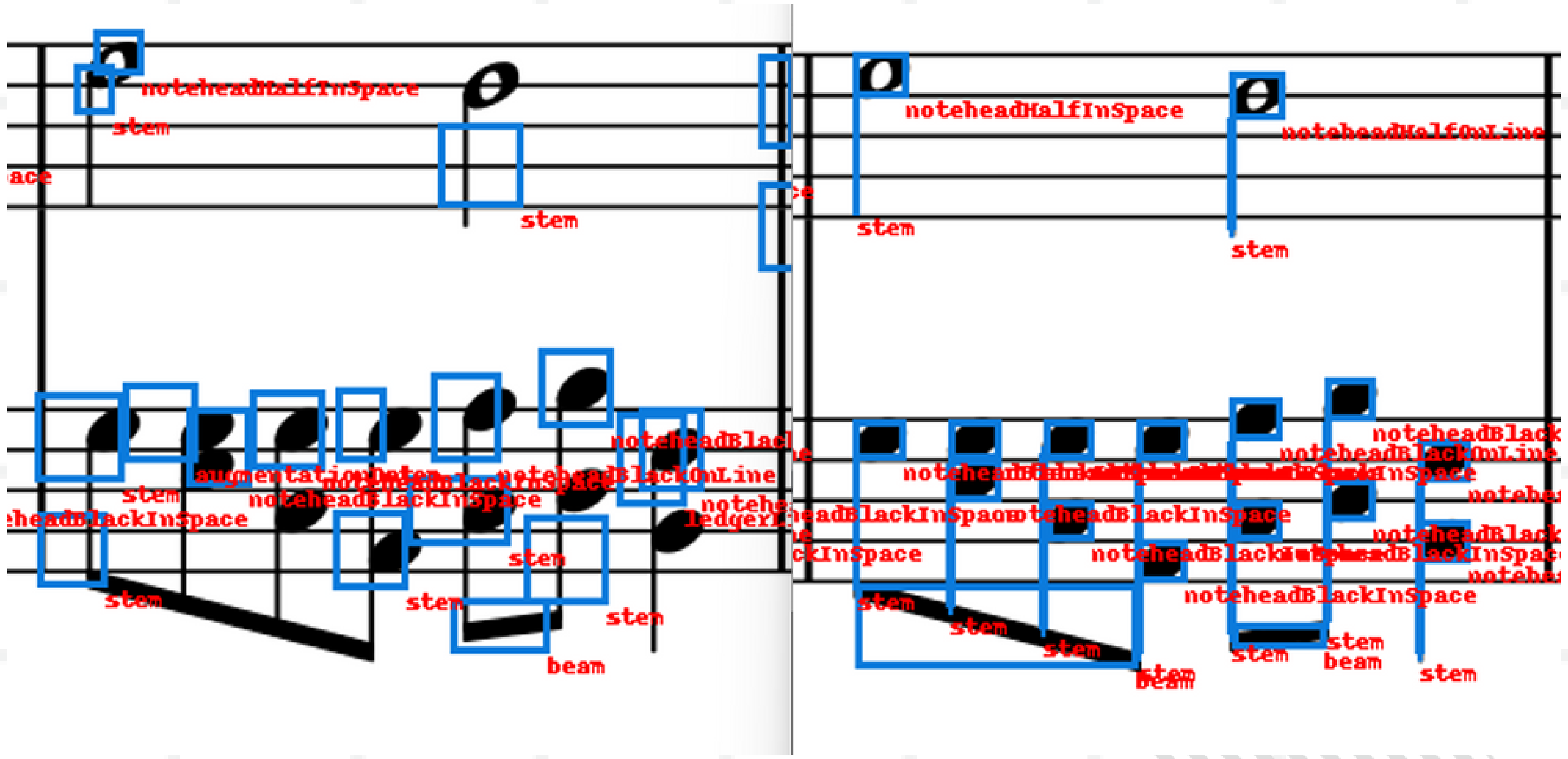
AP at 0.5 : Average precision when IoU is 0.5

DISCUSSION

1. Improved performance

prediction ground truth

Recognize the title of the score as a musical symbol

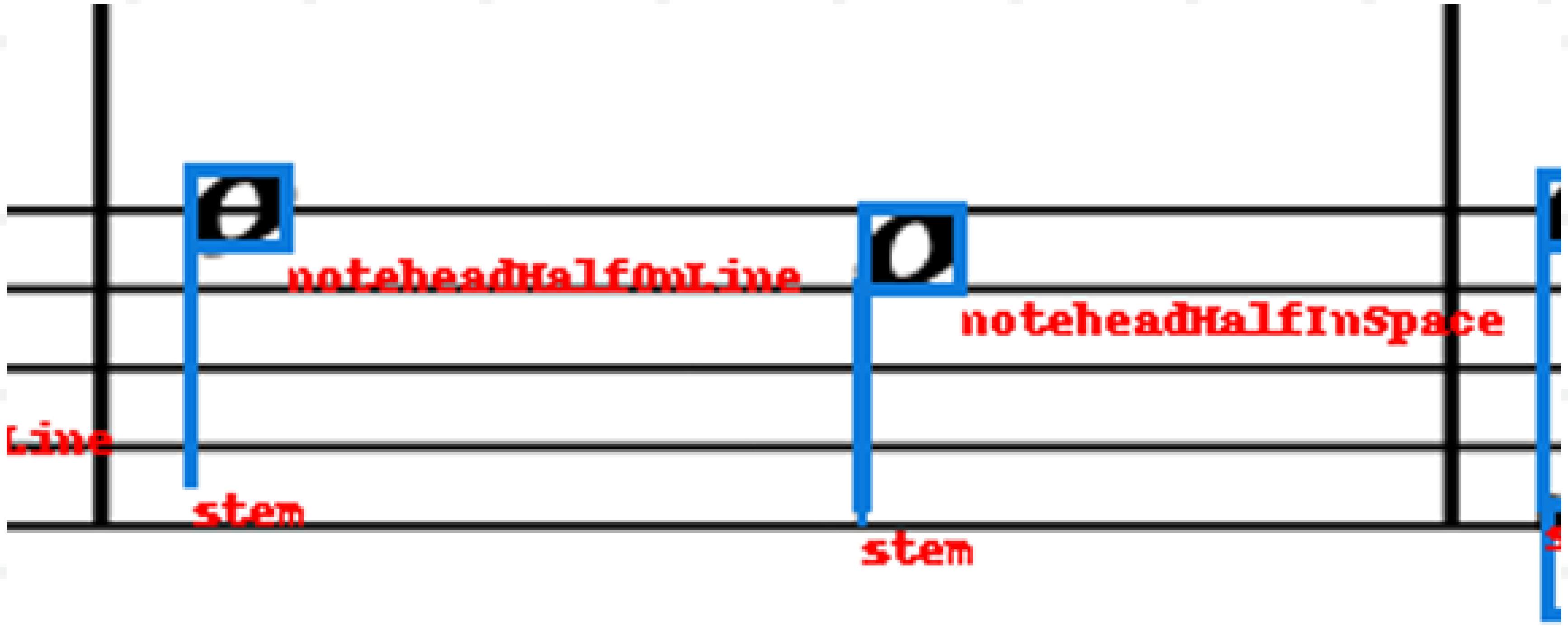


prediction ground truth

Incorrect Bounding Box

DISCUSSION

2. Identifying of syllable name



noteheadHalfInSpace

What can be confirmed with prediction?

- Beat of notes
- Whether the note spans a line
- ~~Syllable name of note~~ ➞ not yet

THANK YOU