Video699: Interconnecting Lecture Recordings with Study Materials <github.com/video699>

Michal Štefánik <stefanik.m@mail.muni.cz>, Vít Novotný <witiko@mail.muni.cz>
MIR research group <mir.fi.muni.cz>, Faculty of Informatics, Masaryk University, Brno, Czechia

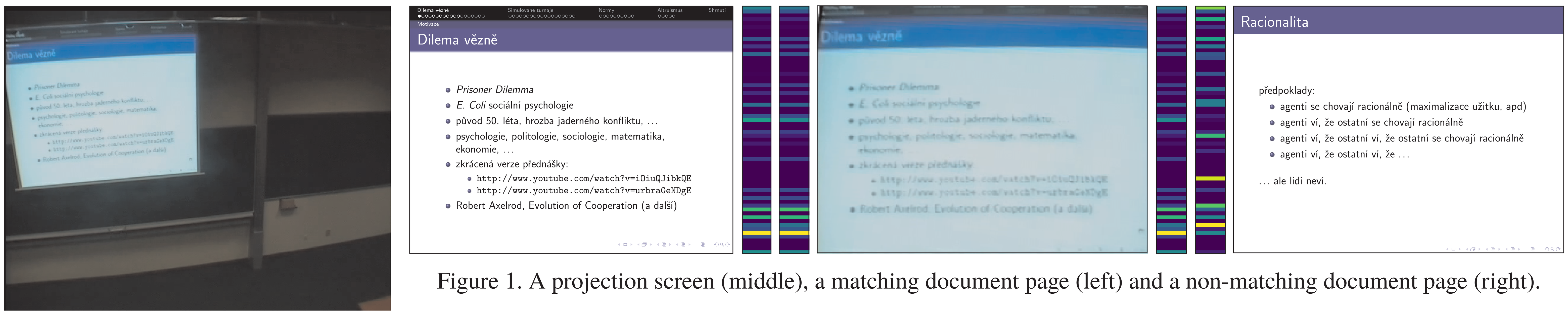

Figure 1. A projection screen (middle), a matching document page (left) and a non-matching document page (right).

Introduction

- Recording lectures is common in academia and massive online courseware.
- Machine-readable information about lecture slides is rarely preserved.
- We map lecture recordings to lecture materials.
- This enables full-text search, lecture recommendation, and video upscaling.

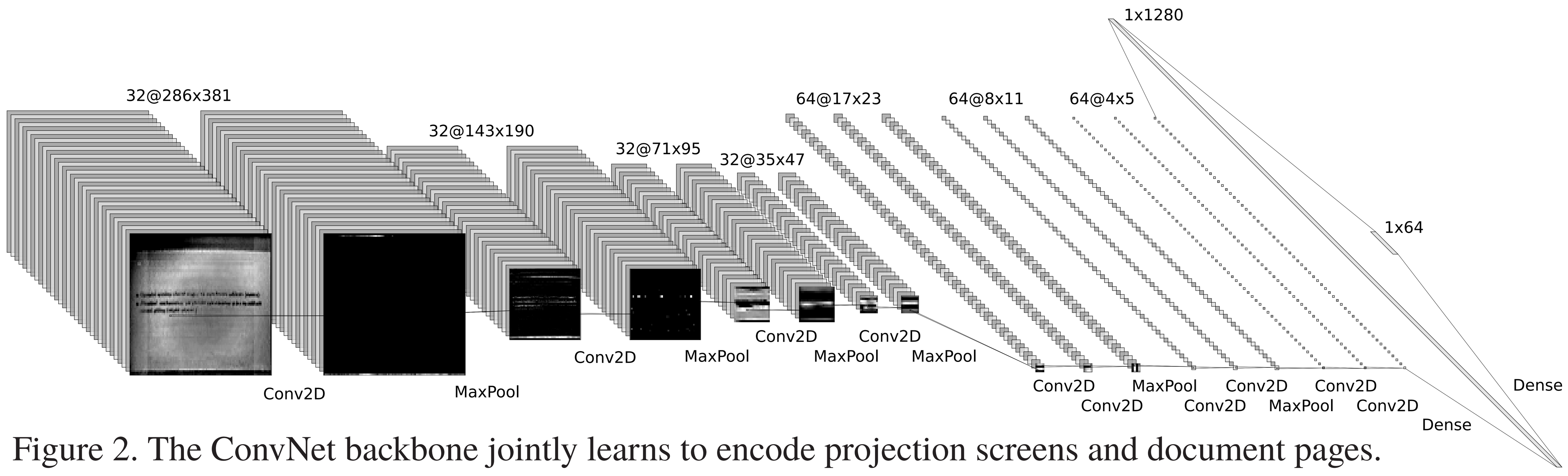

Figure 2. The ConvNet backbone jointly learns to encode projection screens and document pages.

Challenges

- Recorded lecture slides are small, noisy, possibly obscured, optically dictorted, and unevenly-lit.
- Many lecture slides are highly similar.
- We need to scale to large image databases.

Dataset

- We collected a random sample of 17 lecture recordings from 2010–2016.
- We drew a stratified sample of up to 25 video frames from each recording.
- In each frame, we annotated lit projection screens and their condition.
- For each screen, we annotated document pages of lecture materials shown in the screen.

Evaluation

- Estimate classification accuracy by guessing whether pairs of screens and pages match.
- Estimate pixel-level precision and recall by detecting screens in video frames.

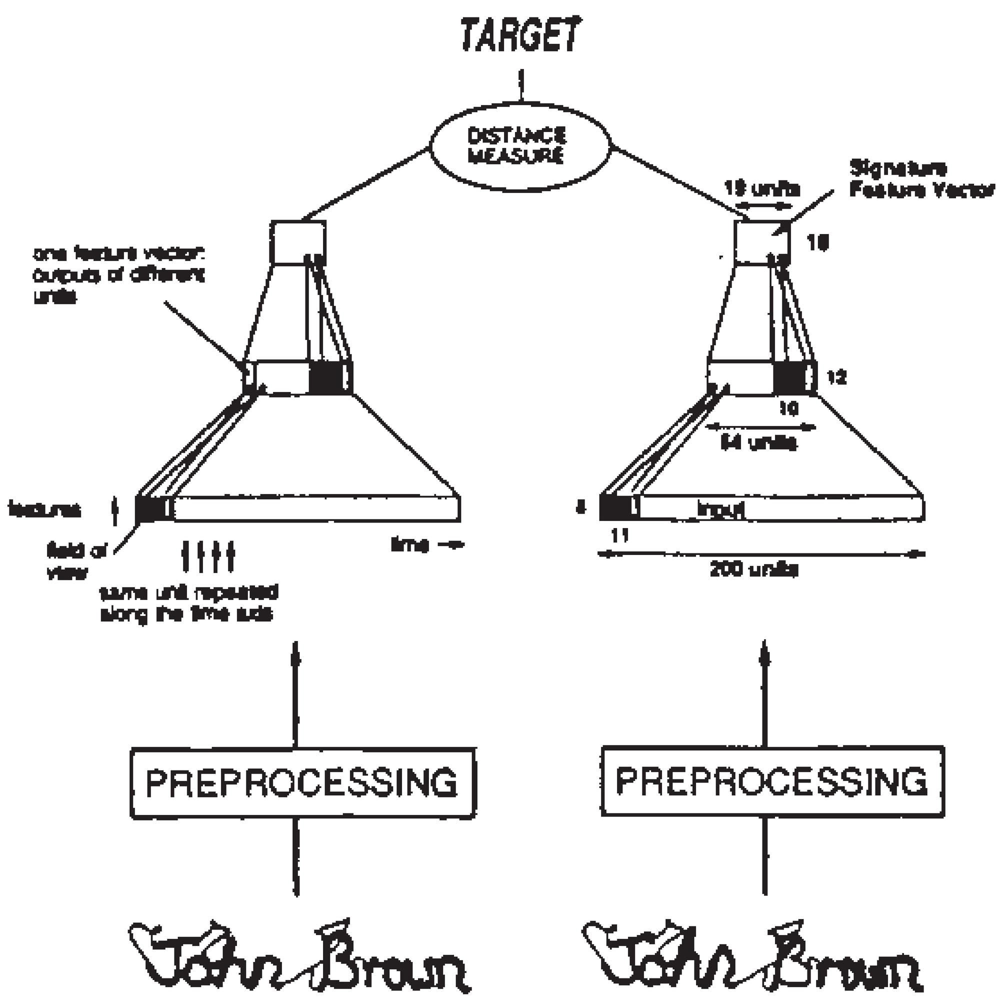

Figure 3. Siamese network [1] classifies image pairs.

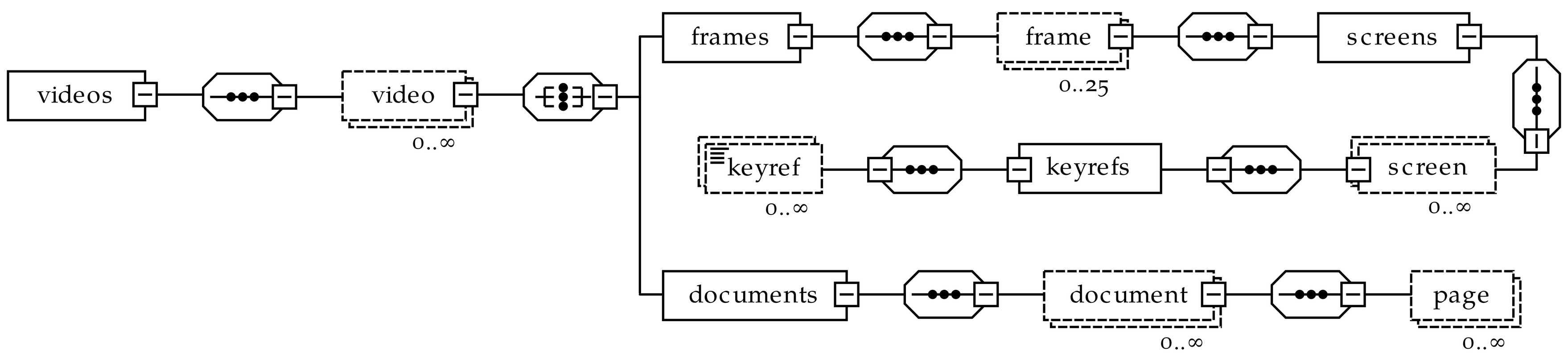

Figure 3. The XML schema of our open-licensed dataset.

Implementation

- We encode lecture materials using ConvNet.
- We detect lit projection screens in a video frame using the SharpMask architecture.
- We classify the nearest document page as a match / no match using a Siamese network.

Results

- We achieved 91% [±5%] weighted validation accuracy using leave-one-out cross-validation on the 17 annotated lecture recordings.
- We correctly classified screens on 70% [±4%] of video frames using leave-one-out cross-validation on the 17 annotated lecture recordings.

Conclusion

- We have developed a production-grade system for mapping lecture recordings to lecture materials.
- We released a richly-annotated dataset under the ODC Open Database License.
- We evaluated a novel method for content-based image retrieval and image pair classification using modern digital image processing.

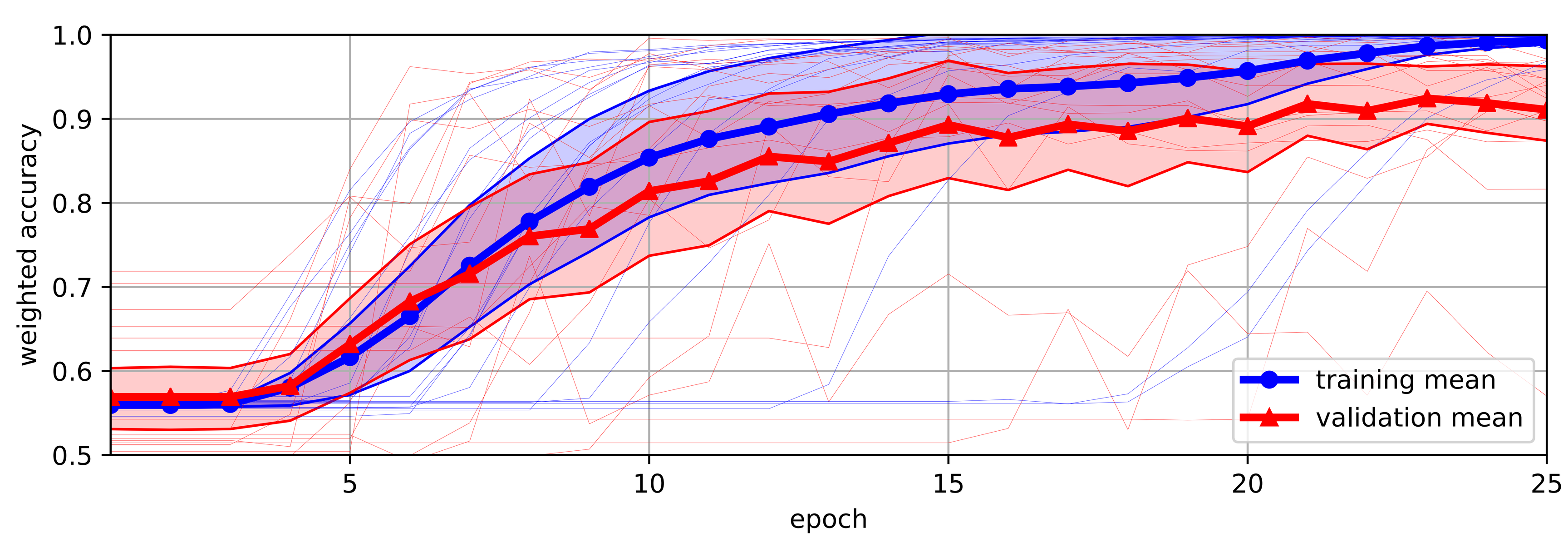

Figure 5. 95% bootstrap confidence intervals of the mean weighted accuracy of our Siamese classifier.

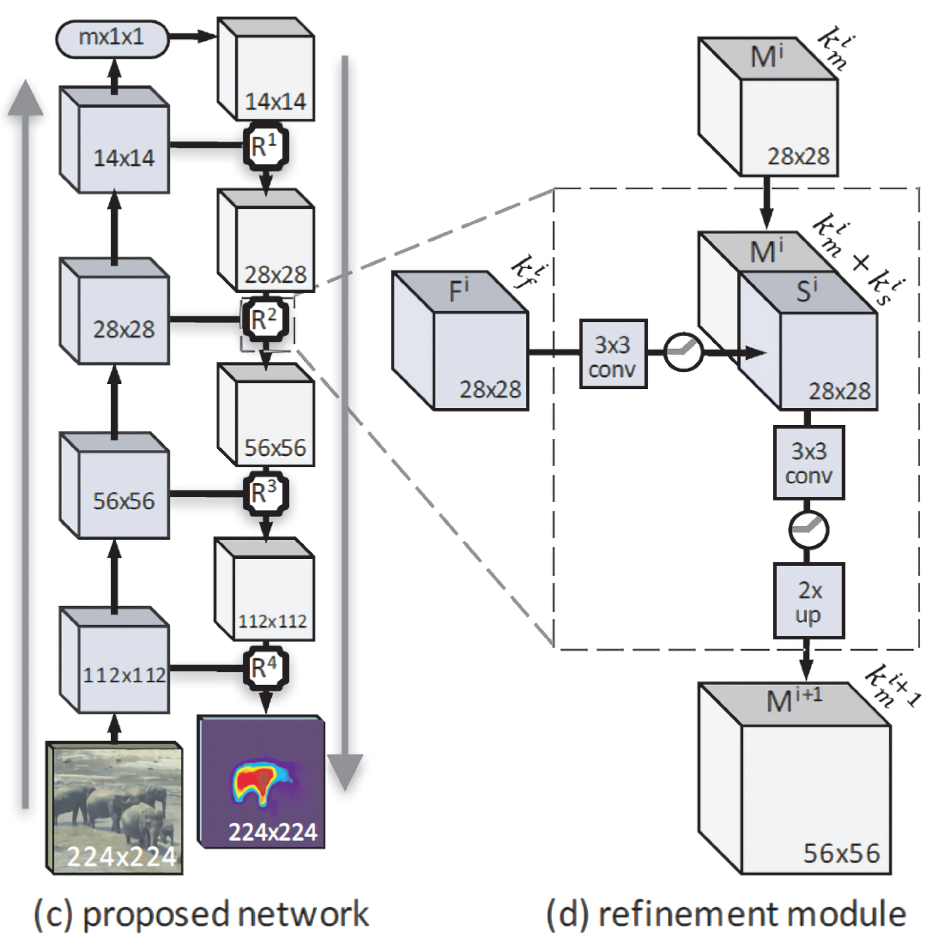

Figure 4. The SharpMask architecture [2].

References

[1] BROMLEY, Jane, et al. Signature Verification Using a “Siamese” Time Delay Neural Network. In: Advances in neural information processing systems. 1994. p. 737–744.

[2] PINHEIRO, Pedro O., et al. Learning to Refine Object Segments. In: European Conference on Computer Vision. Springer, Cham, 2016. p. 75–91.