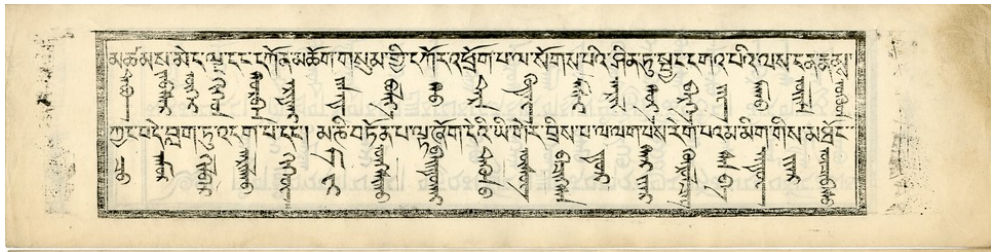
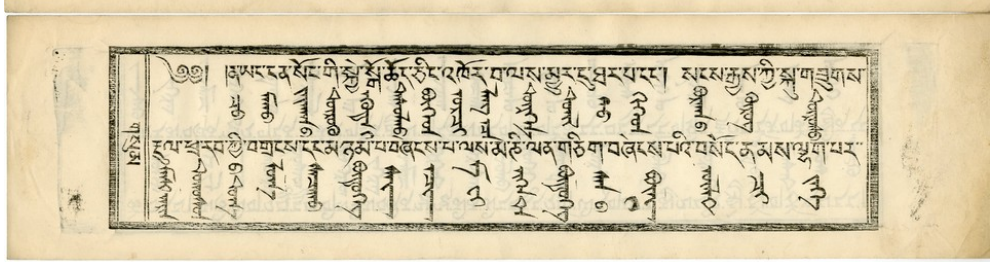
Deep learning for the development of an OCR for old Tibetan books

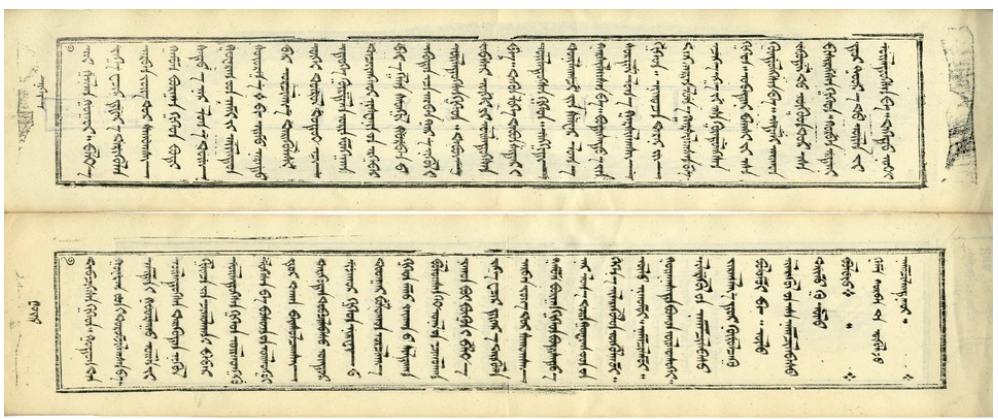
Brodt K.[1,2], Rinchinov O.[3], Bazarov A.[3], Okunev A.[2,4]
[1] Université de Montréal, Montréal, Canada
[2] Novosibirsk State University, Novosibirsk, Russia
[3] Institute for Mongolian, Buddhist and Tibetan Studies, SB RAS, Ulan-Ude, Russia
[4] MTS AI LLC, Novosibirsk, Russia

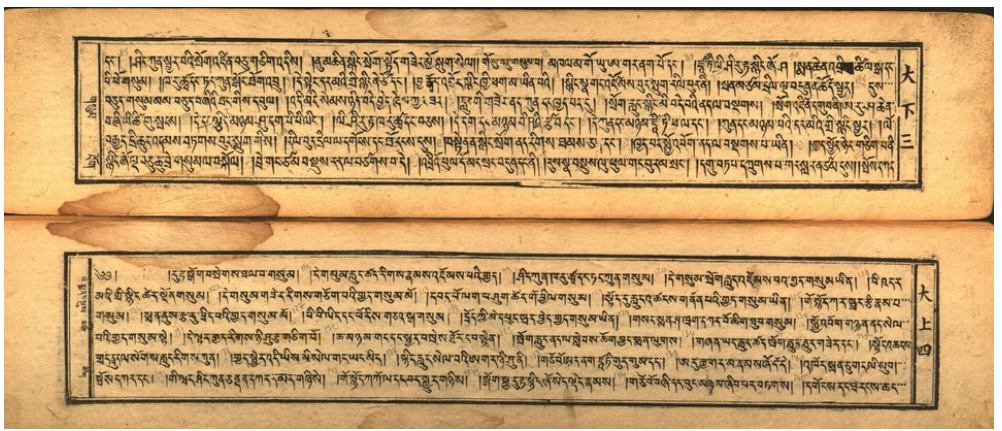
BGRS/SB-2022

Motivation

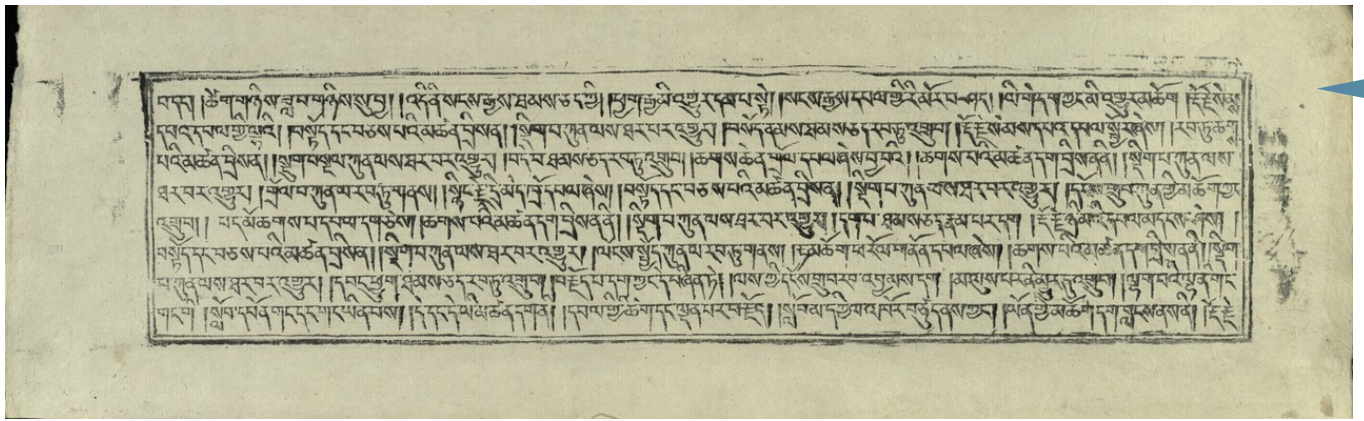






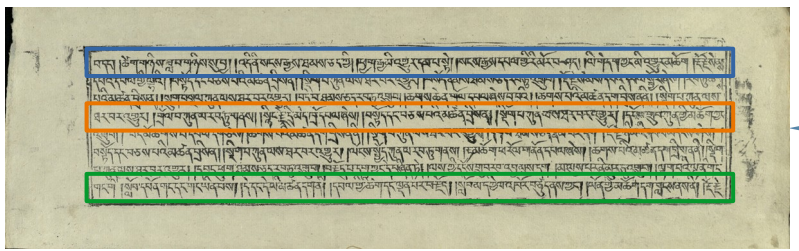


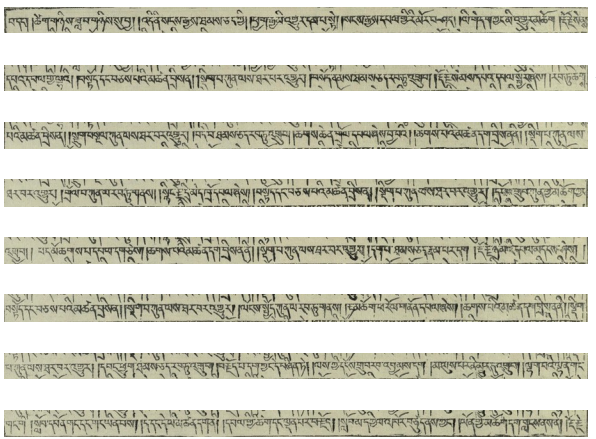
Objective



pa kun las thar par 'gyur//dbang phyug thams cad rab tu 'grub//brjod pa dag kyang de bzhin te/ las kyi dngos grub rab 'byams dag /ma lus par ni myur du 'grub//lhag pa'i lha ni gang

Pipeline

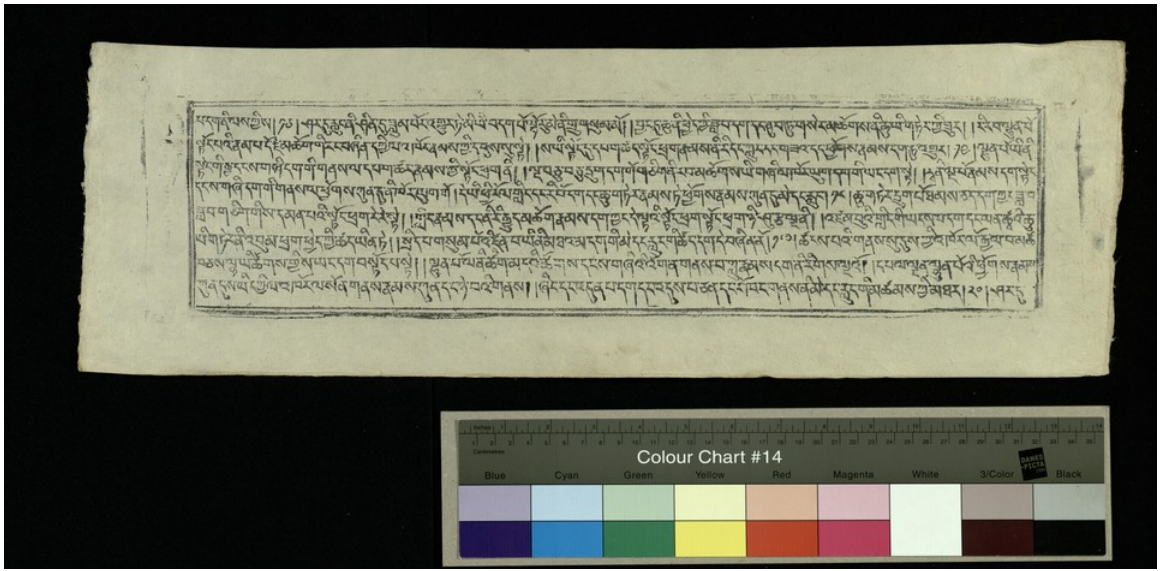




pa kun las thar par 'gyur//dbang phyug thams cad rab tu 'grub//brjod pa dag kyang de bzhin te/ las kyi dngos grub rab 'byams dag /ma lus par ni myur du 'grub//lhag pa'i lha ni gang

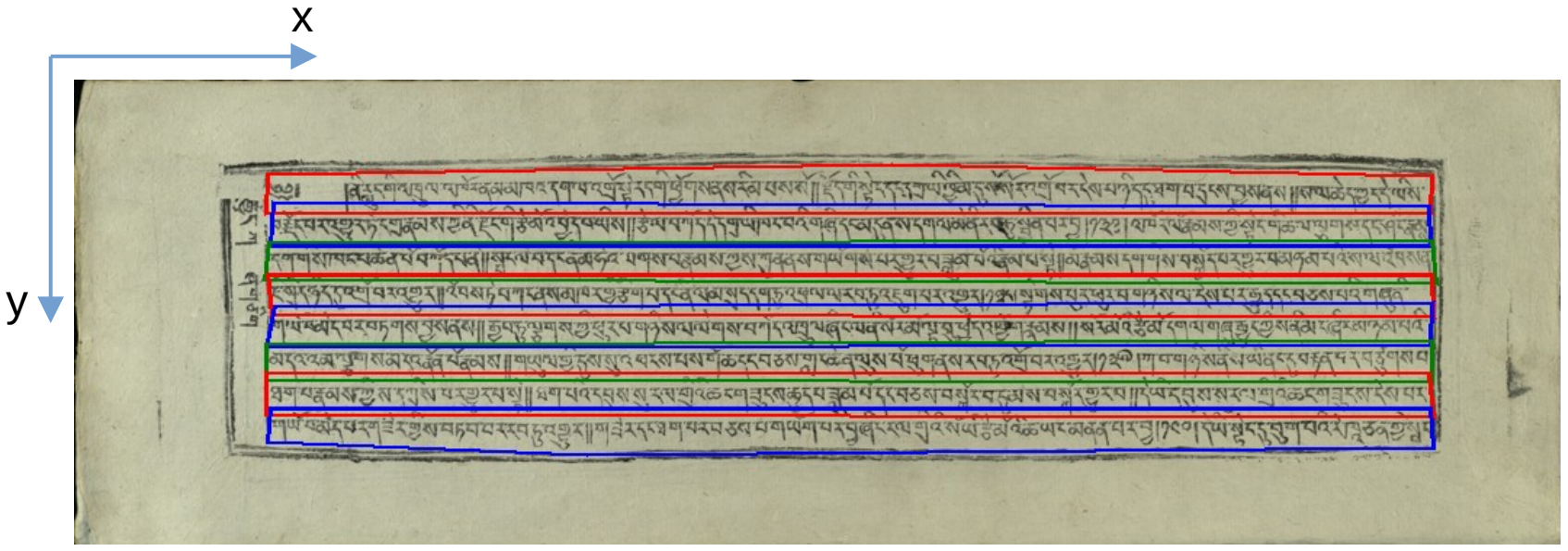
Dataset

- 420 train images, 30 test images
- image dimensions WxH: 2048x650
- 8 lines per image in average



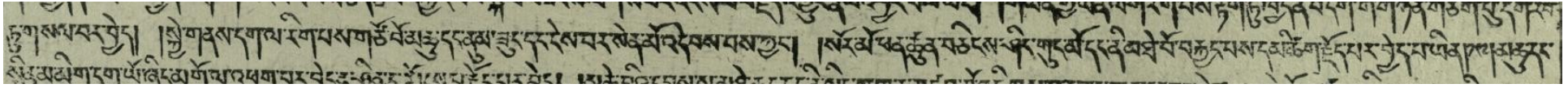
Data: Object Detection

Bounding box: [x, y, width, height]



Data: Optical Character Recognition

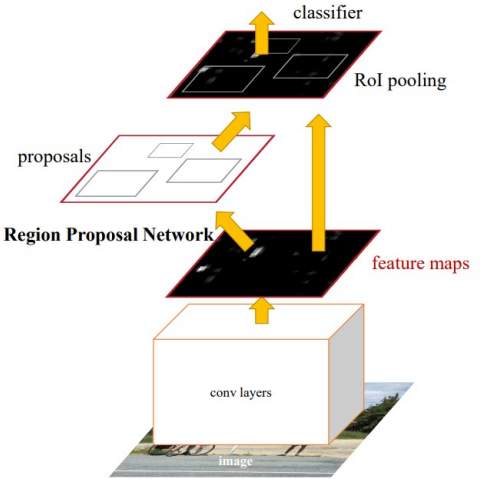
- 50 symbols in Latin alphabet
- 300 characters per line in average
- 3357 train lines, 240 test lines



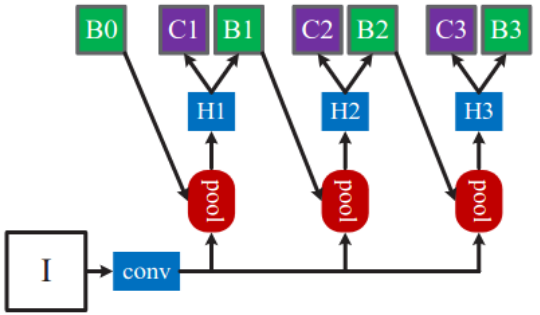
tu gsal bar byed//skye gnas dag la reg pas gtso bo mchu dang nu ma zung dang gnes par sen mo 'debs pas kyang//sor mo sor mo phan tshun bcings shing gung mo dang ni mthe bong brkyang pas dam tshig brjod par byed pa yin/ /mchu dang

Object Detection

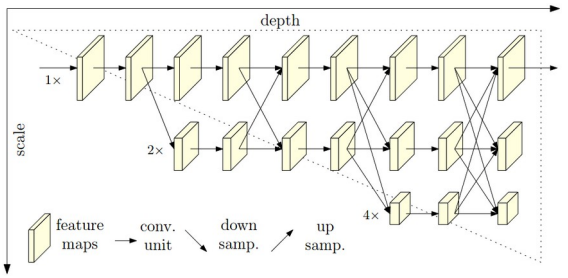
- 2-stage models
- Cascade R-CNN HRNet
- 1 epoch / 2 images per GPU
- 15 min on 8xNVIDIA V100 GPU 32Gb VRAM
- mAP 0.979



[Faster-RCNN, Ren et al. 2015]

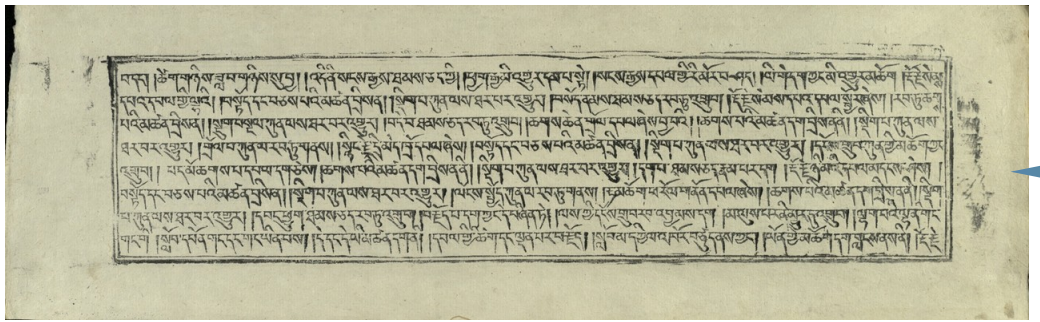


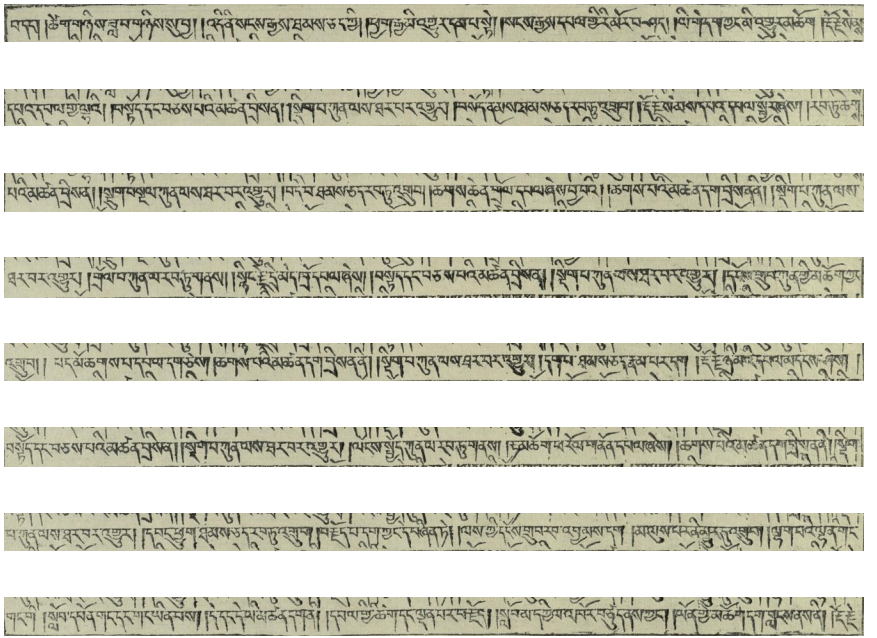
[Cascade R-CNN, Cai et al. 2019]



[HRNet, Sun et al. 2019]

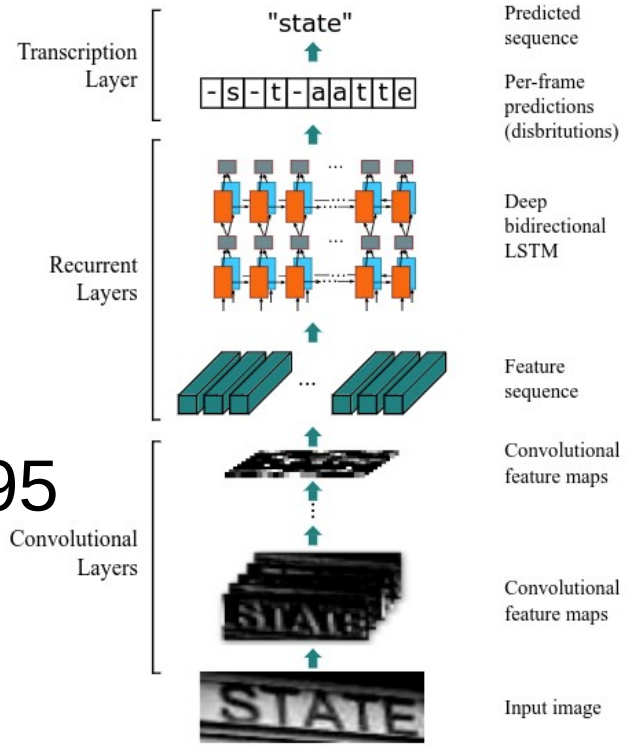
Object Detection





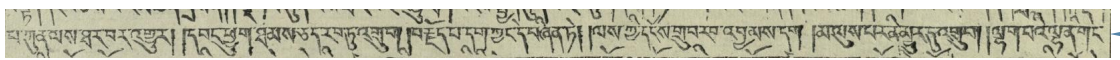
Optical Character Recognition

- CRNN
- 20 epochs / 16 images per GPU
- 2 hours on 8xNVIDIA V100 GPU 32Gb VRAM
- character precision 0.95, recall 0.95
- normalized edit distance 0.94



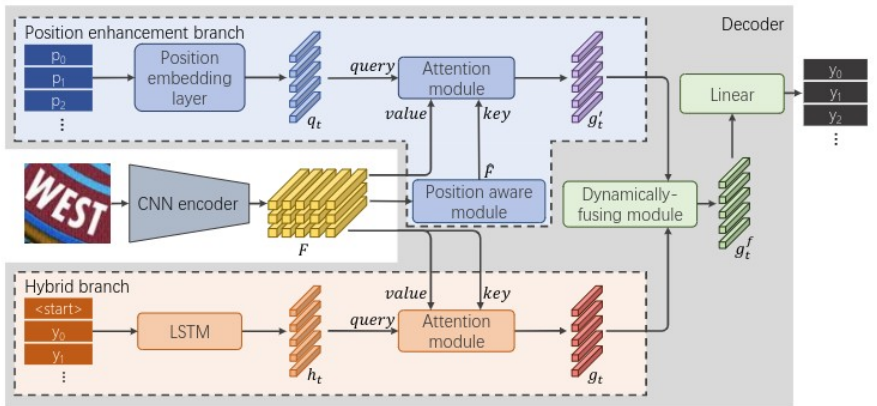
[Shi et al. 2015]

Optical Character Recognition

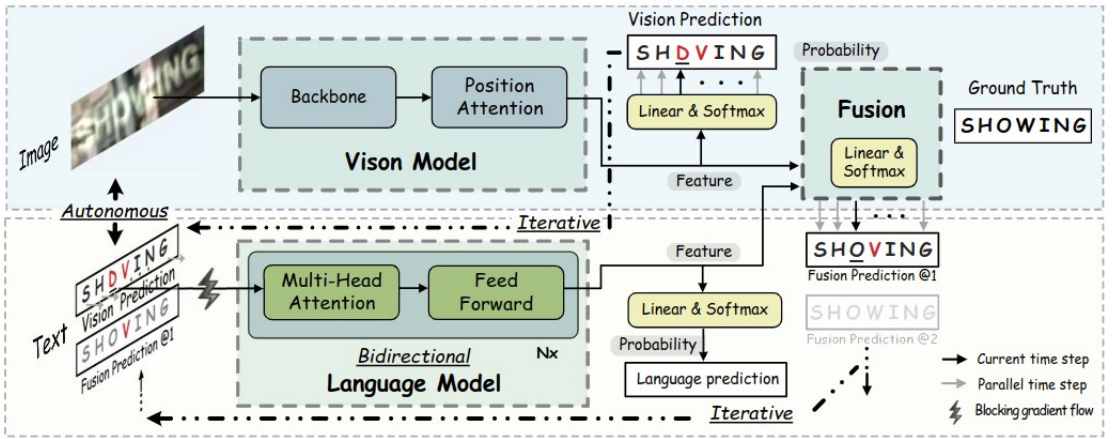


pa kun las thar par 'gyur//dbang phyug thams cad rab tu 'grub//brjod pa dag kyang de bzhin te/ las kyi dngos grub rab 'byams dag /ma lus par ni myur du 'grub//lhag pa'i lha ni gang

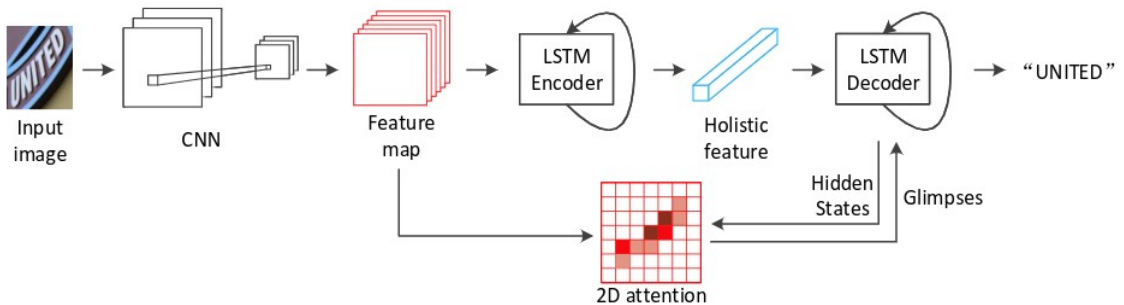
OCR: didn’t work



[RobustScanner, Yue et al. 2020]



[ABINet, Fang et al. 2021]



[SAR, Li et al. 2019]

Conclusion

- full-featured system for OCR of the Tibetan script
  - stream decoding the scanned text in Tibetan
- normalized edit distance 0.94