

## Addition to Last Week's Proposed Pipeline

- We took the previously proposed layout of a system and looked at its edges -> meaning -> what could be done to improve the draft even further?
- The proposed all-including ingredients to make up a processing pipeline for text recognition:
  - Image Scaler
  - Feature Extraction Unit (locate text)
  - Integral Embedding Extractor (group text patches by visual and contextual features)
  - Contextual Text Block Generator (text patch arrangement within their group)
  - Character Classification Unit (real deal character detection within patch group)
- We looked at techniques with the potential for boosting the processing efficiency, as the main restricting factor remains „running on a smartphone“
- The Image Scaler bugged us, we thought this could be made obsolete
- Turns out: **It could not be made obsolete**
- Feature Extraction Unit, right after the Image Scaler, is a Convolutional Neural Network
- Convolutional Neural Networks can take images of different resolutions as input, within a range (set upper image size boundary)
- Literature claims now that having differently sized images vastly decreases the systems ability to focus on the actual image contents
- The same AI that on fixed size images could make out details such as text, with variable image resolution can only make out general image contents with certainty (e.g. image contains a dog or no dog)

At the end of the research: We basically looped back to where we started. But now - with this structure that we would need each component of - what is there to consider if we put it into action on a smartphone?

- as latency is a key factor, processing can't rely on data connections, the smartphone has to process the images itself, not servers in the cloud
- There are two dimensions to this: Memory and battery
- Battery life is linked directly to processing complexity
- The processing complexity can actually be actively reduced
- There are techniques applicable to the proposed pipeline, like Pruning or Knowledge distillation

- The latter actually reduces complexity and increases accuracy, the introductory paper to this concept achieved 98.6% accuracy on a small model which profited from a big trained model's experiences

We seeked to criticize the pipeline proposed last week, but the possibilities and configurability for running on a phone with decent accuracy lead us to still conclude with this proposal, unchanged from last week

### **Additional Task: Competitor Analysis**

- new page in ClickUp assigned:
  - Pipeline made customer think: How do they work (functionally/technologically)?  
What do they **not** do?

### **Task**

- **Find Competitors** -> decide which to investigate (one per student)
- **Describe the user flow** from login to translation to usability of resulting translated text
- **Describe the tech-stack**, what does it look like, what did they put together and why?

### **Goal**

Understand in detail what the **one** selected competitor **per student** does/does good/does poorly

- > Functionality
- > Applied Technologies
- > User Flow
- > Pros/Cons/Learnings for Translue