

# Final Project Proposal

**Team name:** *A Fine Matrix*

## Proposal Questions

### What is your project idea?

Our idea is to re-create a scoped-down version of the Google Lens Translate feature. The end product will be able to take in an image and replace text in the source language with text in the target language.

### What is the socio-historical context that this project lives in?

Google Lens allows people to understand unfamiliar content when in a foreign country. This is useful for traveling, but also for individuals who move to countries where they are not fluent in the primary language.

### Please list three stakeholders that your project could impact, and describe how it could impact them.

One stakeholder is travelers, as mentioned above. They will be able to travel further off the beaten, touristy path and enjoy more rural / authentic trips.

Another stakeholder is potentially interpreters/travel guides. With this new technology, Google has begun to lessen the need for these industries.

Finally, researchers attempting to read a resource only available in a foreign language will be affected. This tool has the potential to enable more cross-cultural research. Though, if it dis-incentivizes actually learning the target language, important linguistic context could be lost.

### What are the skills of the team members? Conduct a skill assessment!

All team members are proficient in python and general computer vision techniques. We have two team members currently taking the Deep Learning course, which may prove handy in training the CNNs needed. Additionally, all team members are proficient in the reading of a foreign language, allowing us to better grok how well the final project performs.

**What data will you use?**

We're planning on using natural-scene data from the 2019 ICDAR Multi-Lingual Scene Text [competition](#). It includes ten languages, but we may use only a subset of these languages to facilitate training.

**What software/hardware will you use?**

We plan to use either the CS Grid or Oscar for compute as we foresee that training will require significant resources. We will use the [PyTorch](#) python library to train our CNNs. We will not re-create a machine translation model from scratch, instead we will use the Google Translate API for the actual translation.

**Who will do what?**

All team members will discuss overall approaches and make design decisions together. All team members will work on the architecture of the CNNs.

Periphery Tasks: - Team member 1 will validate and pre-process the dataset. - Team member 2 will implement the graphical interface (inputs, etc.) - Team member 3 will hook up the google translate API and re-draw translations onto the bounding boxes surfaced by the detection network

**How will you know whether you have made progress? What will you measure?**

We will have made progress if the end-to-end model produces meaningful results on images of increasing difficulty (well-lit and level → poor illumination and skewed). We will measure accuracy and other metrics on a test set.

**What technical problems do you foresee or have?**

We foresee that computation may take a significant amount of resources. We believe that errors may accumulate throughout the models (if detection is slightly off, recognition will prove more challenging).

**Is there anything that we can do to help? E.G., resources, equipment.**

We're excited to talk with our mentor TA and receive general feedback, but feel confident as of now!