Second Sight

W210 Synthetic Capstone

Camille Church | Amanda Teschko | Cameron Wright | Lisa Wu

The Team

Camille Church

Amanda Teschko

Cameron Wright

Lisa Wu

Problem Space

In the US alone,
20 million
people are affected by visual impairments

Of those affected,
27%
are living below poverty level

Prices can reach
$5000+
for other tools on the market

Our MVP Solution

Solution
Our solution combines text image recognition models and text-to-speech capabilities, in an application that is both affordable and safe

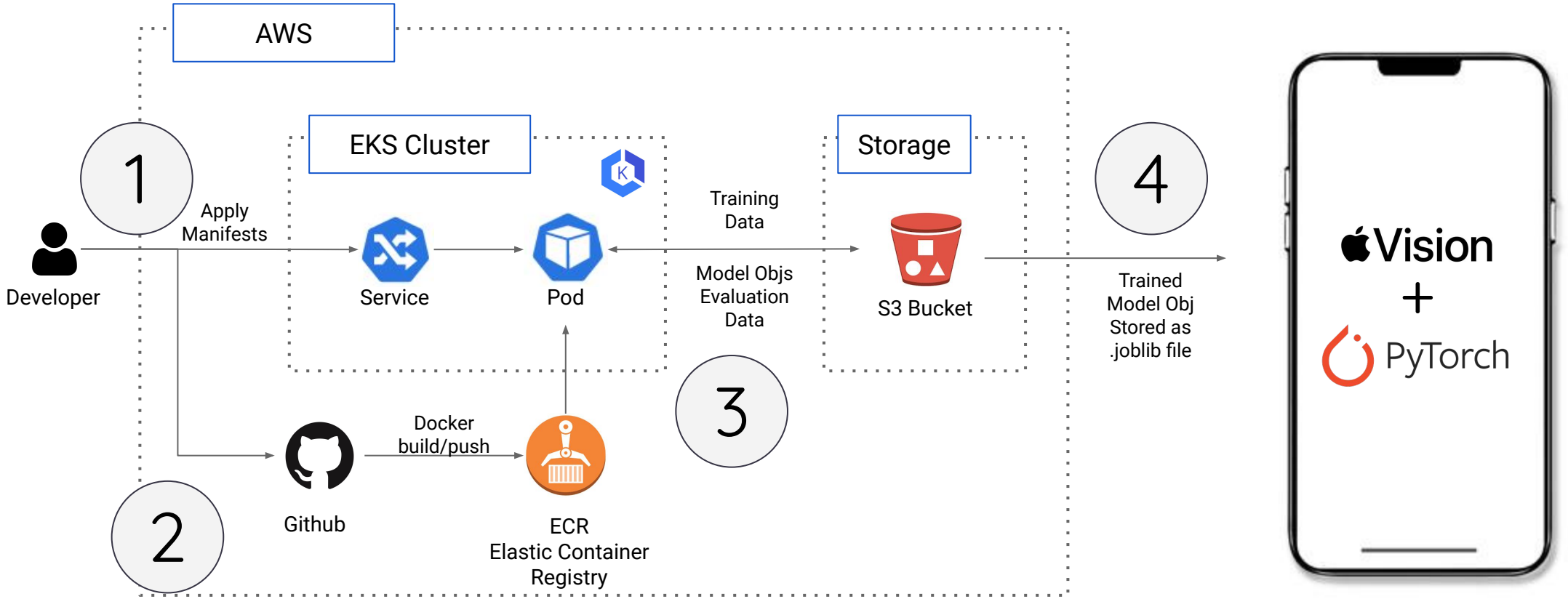
Impact
Our project endeavors to provide a mobile application that is free to access for visually impaired people to aid their day to day lives

Target User
Those struggling with reading text (Presbyopia), who seek more affordable care, and own smartphones

User Need		Feature
Ability to understand handwritten and typed text	→	Fine tuning typed text OCR models to understand handwriting
Easy to open and use app	→	Voice enabled prompts and a simple app interface
Confidence and trust in the results	→	Warnings to indicate when results are not clear or caution should be exercised
Cost effective	→	All processing done on phone; Free to use app

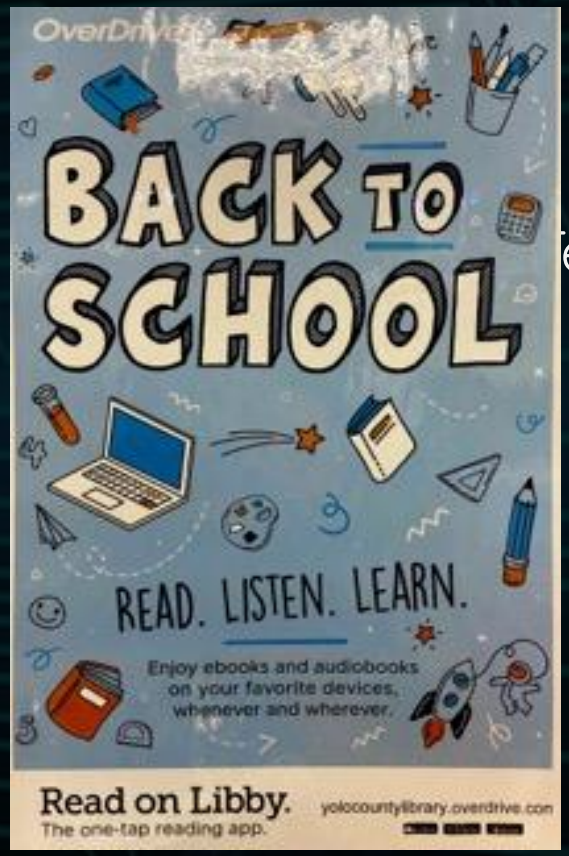
Demonstration

End to End Architecture

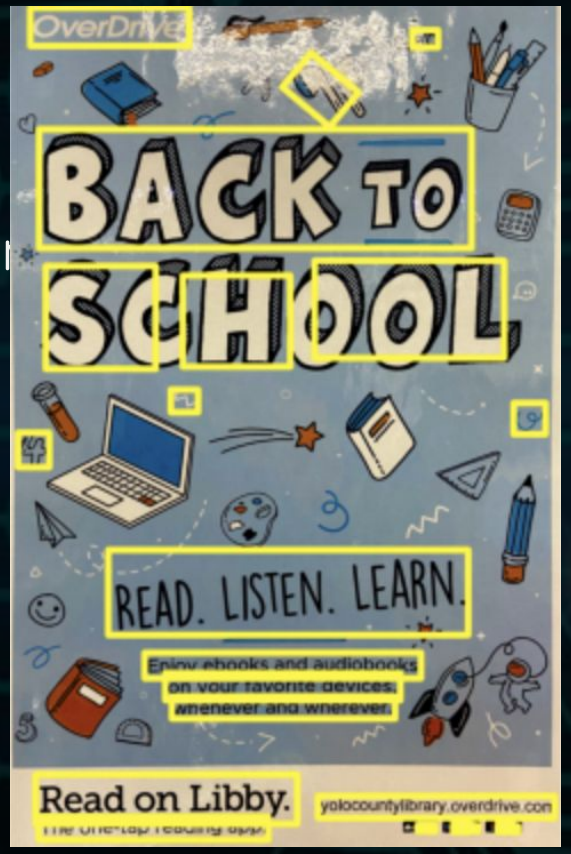


General OCR System

we only address Text Recognition in our modeling



Text Detection



Text Recognition

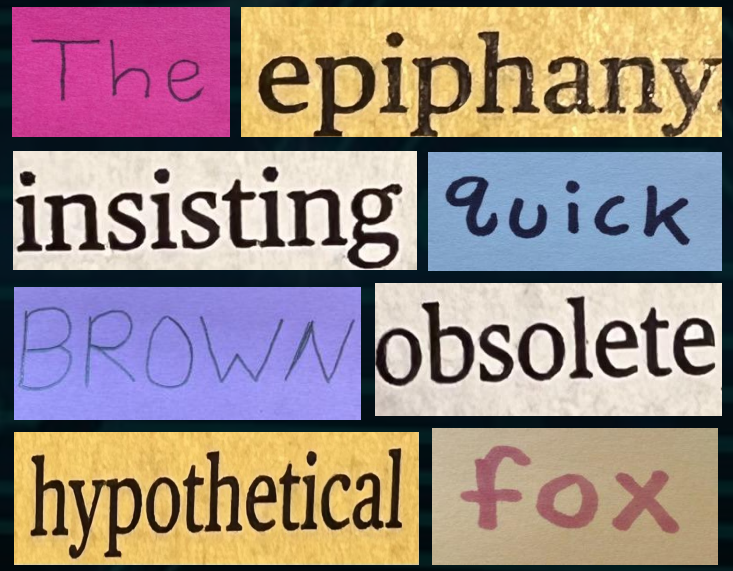
```
['OverDrve',
 'BACK 7',
 'S@HOoL',
 '(9',
 'Read. L/Sten . LEARn.',
 'Enjoy ebooks and audiobooks',
 'on your favorite devices,',
 'whenever and wherever.',
 'Read on Libby',
 'yolocountylibrary overdrive.con',
 'canoy',
 'The one-tap reading app.',
 'p Store',
 'Googkc Play',
 'Microsoft',
 'C']
```

Modeling Approach

Pre-trained Transformer (TrOCR)

Hugging Face

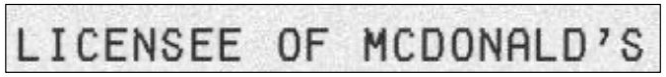
Task Specific Datasets



we developed our Text Recognition model by fine-tuning a pre-trained transformer on task specific datasets

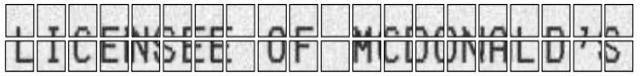
Model Architecture (TrOCR)

Step 1: Raw Pixels



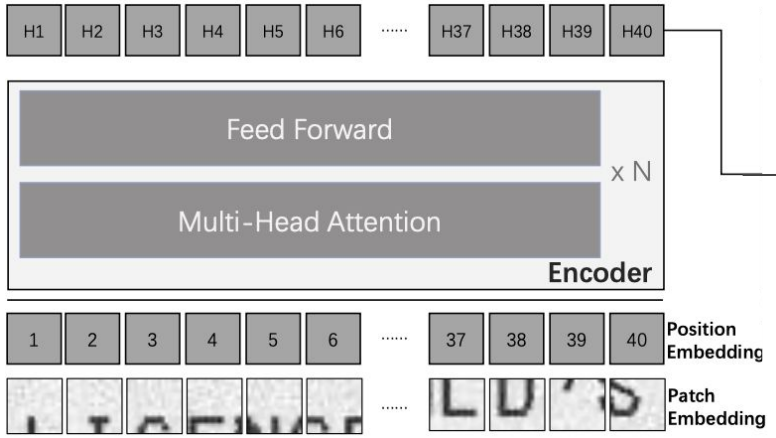
- *Read in the raw pixels*

Step 2: Image Patches



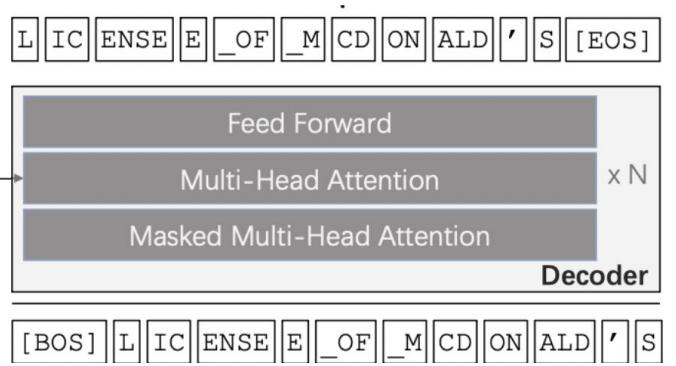
- *Resize and segment the image*

Step 3: Hidden States



- *Create patch embeddings and pass these through encoder layers*

Step 4: Word Pieces



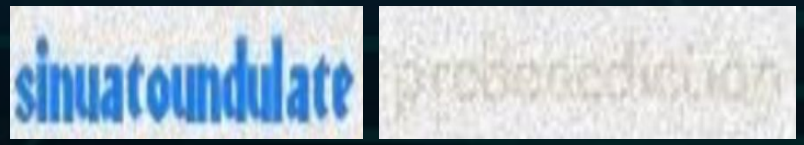
- *Produce words using hidden states and previous word predictions*

Source TrOCR: Transformer-based Optical Character Recognition with Pre-trained Models

Model Experiments

TrOCR Model Version	Model Training Work
Base Handwritten (HW) Model	no fine-tuning
Large HW Stage 1 Model	fine-tuned for 20 epochs
Large Printed Model	no fine-tuning

We used both handwritten and printed datasets in our model experiments

Dataset	Text Image Type	Key Challenges	Image Examples
IAM[1]	**Handwritten**	• Unique handwriting styles • Variety in text length • Variety in image quality	
Font Recognition Dataset [2]	**Printed (Uncommon Words)**	• Variety in text length • Variety in image quality • Uncommon words	
Self Dataset [3]	**Printed (Common Words)**	• Variety in text length • Variety in image quality	

1. IAM is a handwritten dataset with 96k text images in English (link)
2. Font Recognition dataset is an unlabeled dataset obtained from Kaggle (link). We manually labeled ~1400 images
3. Self dataset is a new dataset (~ 300 text images) that we generated by taking photos of various text images.

Model Evaluation Metrics

- **Measure Accuracy by individual letter**
  - Developed by our team
  - Two versions: case sensitive and case insensitive
- **Measure Accuracy by the entire word**
  - Developed by our team
  - Two versions: case sensitive and case insensitive
- **Measure Character Error Rate (CER)**
  - CER is common metric for text image recognition
  - The lower the CER value, the better the performance of the model

Key Challenges & MVP Solutions

- **Handwritten text images are very challenging to model and predict**
  - Unstructured data
  - High variance in style, image quality and text length
  - Order matters for sequential data
- **A large size internal model(s) is difficult to house on a mobile application**
- **MVP Solutions:**
  - Fine-tuned TrOCR models with the IAM dataset
  - Used Apple OCR model in our mobile application

Model Evaluation Results - Word Accuracy

	Handwritten Text Data	Printed Text Data	
Model	IAM	Font Recognition Dataset	Self Dataset
Base HW Model	65.2%	n/a	n/a
Large HW Stage 1 Model (Fine-tuned)	68.8%	33.7%	32.7%
Large HW Stage 1 (Fine-tuned + NLP Spell Check)	72%		53.3%
Large Printed Model		84%	93%

Decent Model Performance

Model Performance near the Target State

Key Learnings

- Typed and handwritten text may require separate models
- Keeping processing on the phone will limit how complex and large our model can be
- Pre-trained transformers can provide a great start when developing OCR models
- Finding high quality datasets for fine tuning is challenging -> may require creating your own

Key Learnings & Product Roadmap

- **[Model]** Progress generalizability of model and conduct end-to-end testing for challenging use cases (such as curved or blurred text and currency)
- **[Model]** Add non-English languages to our model development
- **[Model]** Consider opportunities for model size reduction or explore implications of web based solutions

- **[App]** Launch to a broader community of users for more testing
- **[App]** Incorporate features from user feedback
- **[App]** Point application to our latest model


2023-12-02 10:30:59

Our Mission

Democratize text recognition technologies and aid a population of people worldwide who are struggling with vision impairment or weaknesses

Join our mission by downloading **Second Sight** *today!*

Questions

Download our App