Twitter Thread by Alex Strick van Linschoten



Alex Strick van Linschoten
@strickvl



I recently switched what I spend the majority of my professional life doing: history -> software engineering. I'm currently working as an ML Engineer @zenml_io and really enjoying this new world of #MLOps, filled as it is with challenges and opportunities.

I wanted to get some context for the wider work of a data scientist to help me appreciate the problem we are trying to address @zenml_io, so looked around for a juicy machine learning problem to work on as a longer project.

I was also encouraged by @jeremyphoward's advice to "build one project and make it great" (https://t.co/Doo88EUhkN). This approach seems like it has really paid off for those who've studied the @fastdotai course and I wanted to really go deep on something myself.

Following some previous success working with @adyantalamadhya and another mentor via @SharpestMindsAI on a previous project, I settled on computer vision and was lucky to find @ai_fast_track to mentor me through the work. (We meet a couple of times per week).

In the last 6 weeks, I've made what feels like good progress on the problem. This image offers an overview of the pieces I've been working on, to the point where the 'solution' to my original problem feels on the verge of being practically within reach.



After just a few lessons of the @fastai course, I trained a classification model to ~95% accuracy to help me sort redacted images from unredacted images.
https://t.co/E0WY4xYnlZ

I used @explosion_ai's Prodigy tool to annotate an initial round of data to pass into the next step, enjoying how the labelling process brought me into greater contact with the dataset along the way.
https://t.co/bNd9mx7FHv

I switched to using IceVision (co-created by @ai_fast_track) to help me with the more complicated object detection problem, using MMDetection and VFNet to get pretty good results early on.
https://t.co/EZ3cdkF8I8

I'm currently in the process of creating my own synthetic images to boost the annotations I've manually made. (I'll be writing about this process soon as well, as I'm learning a lot about why this is so important for these kinds of computer vision problems.)

I've also been amazed at the effectiveness of self-training (i.e. using my initial model in my annotation loop to generate an initial set of annotations which I can easily amend as appropriate, then feeding those annotations in to create a better model etc). More to follow...

I started using @EvidentlyAI to do drift detection, inspired by work I was doing for @zenml_io to add it as an integration. This helped me think about how new data was affecting the model and the training cycle. There's a lot of depth here, and am looking forward to diving in.

I made a tiny little demo on @huggingface Spaces to show off the current inference capabilities. This is a simple @Gradio app but I liked how easy this was to put together (a couple of hours, mainly involving some build issues and a dodgy `requirements.txt` file)

Along the way, I found it sometimes quite painful or fiddly to handle the PDF files that are the main data source for the project, so I built my own Python package to handle the hard work. https://t.co/bghhYDgKOz

I used @fastdotai's nbdev to very quickly get the starters of what I'm hoping might be a useful tool for others using PDF data for ML projects.
https://t.co/aZaTgOUu33

Throughout all this, @ai_fast_track has been patiently helping guide me forward. He saved me from going down some dark rabbit holes, from spending too long studying skills and parts of the problem that needed relatively little mastery in order to get to where I am.

Farid (@ai_fast_track) has been a consistently enthusiastic and kind advocate for my work, moreover, and this has really helped me stay the course for this project that takes a decent chunk of my time (esp seeing as I do it completely aside / separately from my day job).

I feel like I'm consistently making progress and learning the skills of someone working in computer vision, even though I have so much left to learn! My project still has a ways to go before it's 'done', but I'm confident that I'll get there with @ai_fast_track's support. /end

(Reposted this in an ever-so-slightly expanded form on the blog -- https://t.co/2d8rQWn3E7 -- for those who prefer to save things like this to @instapaper or @Pocket)