FUTURS FANTASTIQUES

Vendredi 10 décembre 2021

Grand Auditorium

Jolan Wuyts, Annotating datsets for computer vision to recognise architectural and artistic styles: lessons from the V4Design project

Thanks everyone so I'm Jolan Wuyts I'm a collections editor at Europeana and I will be talking a bit about the work that we did during the V4Design project which is a Horizon 2020 project that I was a project manager for, over the past three years. So I'll mostly be talking as a project manager and not really as a machine learning expert. I'm sure there's a lot more expertise in this room than that I have so I'll try to keep it high level. If I have some time left I'll get to what the actual lessons were but I'll just first talk about some of the work that we did.

So V4Design was a project where, with Europeana and with other partners, we basically wanted to get cultural heritage as close as possible to a user group that we hadn't really looked at a lot before that we hadn't really catered to a lot before which were architects and creative designers so people that used 3D models in their own designs that created new 3D environments in software programs such as Unity or Rhino 3D using digital cultural heritage. So the main output of this project was actually a plug-in both for Unity and for Rhino 3D. In this demo you can see the plug-in on the right, there where people could actually basically search through different databases like Europeana like Wikimedia commons and find interesting digitized cultural heritage that they can then really just drag and drop into their 3D modeling program and reuse. So either those are 2D assets that they could reuse and maybe trace over or use as inspiration or they are also already made 3D objects that they could reuse in their project.

One kind of small auxiliary extra task in this project was that we wanted to see if we could enrich the data that we had with metadata that was relevant or interesting to architects and video game designers when they were creating their work. So we did a lot of user testing to see what kind of searches that architects mostly did and the results we got back were pretty surprising. So there's two main things apparently that an architect mainly searches for when they're trying to create something new. First off is they use a lot of isms so architecture terms or art-isms so like rococo, baroque, gothic architecture, things like that. And the second thing they do a lot is they use emotional words or moods a lot to look for the content they're looking for so they would say "oh I want to create a new building that is very depressing and moody and based on gothic architecture".

So the question we had was could you create some visual recognition algorithms that create metadata fields for emotions and for architecture styles? The reaction we got

to that was a dam that's really hard, that's a really complex task mostly because you know all of these things are very subjective emotions and moods of course are the pinnacle of subjectivity, but even tagging something with an architecture style we didn't really have a lot of ground truth tags for just because curators you know are very specific about what to tag and if you give them a church, you know different annotators will use different styles to actually tag that with: a church might have a roman catholic base and then the second story may be gothic inspired and there may be baroque elements and rococo elements all in the same thing. So how can we capture this complexity in ideally some very base simple classes in a visual recognition algorithm was basically the challenge we faced.

So in my examples I'll mostly use the recognizing architectural styles from images just because in 10 minutes I cannot talk about all three of these examples but so the first thing we did through this process was extract a large dataset of images from buildings from both Europeana and from Wikidata and Wikimedia commons and the first task that we asked our annotators was to create basically bounding boxes to differentiate foreground from background. So this is the VGG annotation tool and basically they made sure that we could create masks showing what is relevant to the building and what is not relevant to the building. Then we got some of these nice mask images that show this part is foreground this part is background. And from that we went through a second round of annotation where we on the one hand asked annotators to validate those objects that already had a tag attached to them. So a lot of Wikimedia commons images had Wikidata information that had a label for instance modernist building or gothic building or nouveau building, so we asked them to validate that and for certain items that didn't have any ground truth yet we asked them to add a certain tag to this. So we created a bespoke online really simple annotation website where people could connect to and tag their objects with. And because every artificial intelligence presentation needs at least one confusion matrix, this is ours. Some classes (we ended up with 18 classes) performed really well, others are still very messy you'll see in the bottom left corner there's quite some confusion around neoclassical and these types of classes.

So from this work there's quite a lot of lessons and that we got. The first big one is the need for very clear guidelines especially when it comes to these very subjective classes, of course there will be bias in these kinds of datasets because the subject matter for these classes is inherently subjective. What we wanted to get to was to have a kind of annotation that was as homogeneous as possible in that bias and to homogenize that bias we had very very rigorous annotation guidelines in place. So those guidelines for instance very clearly delineated what every architecture style was like, what the examples where, it gave very clear you know keywords and we also told annotators that a few of the objects that they were tagging already had been validated by us in the background and would basically be used to check if they were following the guidelines that we set out for them or not. So these were kind of gold examples that we could see if they were biased in the right way or not. We also found

that validating an already existing ground truth tag is a lot easier for annotators and creates a lot less heterogeneous bias than creating their own tags so clicking their own class.

Secondly in what data or metadata to show the annotators we also could homogenize that bias a lot more, so instead of showing these annotators both the foreground and the background and the rest of the metadata of an object, we only showed them exactly what the visual recognition algorithm would see as well which is the blacked out foreground masked image. That way annotators couldn't find any other context clues in other metadata tags, for instance if they knew what the creator of a building was or where it was located or when it was built they might change their tag or change their class which the algorithm wouldn't pick upon so this will create other noise in the dataset.

And then lastly this might be very self-evident but agreeing on intellectual property of what you're working on in advance can really help this whole process along so mostly because this was a European Horizon 2020 project near the end of the project there was a lot of talk about how do we make this sustainable, how do we make sure that these plugins stay online and these services stay running after the funding stops. So there were a lot of discussions about partly commercializing these plugins which in and of itself is totally fine but we had agreed from the start of the project that the training and test datasets that we would create would stay completely open, completely open source, freely reusable, also because that was a requirement for us to use certain datasets that we could then use in our training sets. So for instance a lot of our data was tagged as a CC by SA which means it's a sharelike license so you have to re-share everything you used with the same creative commons license or something more open.

So to conclude luckily because of that some of these datasets are now available on Zenodo and I warmly invite you all to check them and maybe use them as a basis for further work in this in my opinion really interesting field.

There we go. Thank you all for listening.