use tensor flow - open source by google

keras - library built upon tensor flow

in tf everything is array of numbers

keras makes it easier

cv2 - computer vision library python

get that installed. processes images

convert color pixels into grayscale. makes into a range

use glob - reads all files in a directory without having to loop - helper library

classification metrics

admissions, linear algebra, stats, discrete math, matrices, python

ideally would want want more images per listing, and don't need as many listings

Project Notes:

Originally wanted to make it not only identify what was a good flipping opportunity but *also* train it to figure out *how much* you'd make from it from flipping and from renting

But the thing is was told to ONLY do images OR text - *NOT BOTH!* cuz that'll get pretty hard for now. So I wanted to do something different, so chose only images, as a good learning experience. but then the issue is without text data like sq ft, year built, estimated SELLING COST & *RE-*SELLING COST, I couldn't do the initial idea. I could only do basic classification. so then that means I have to manually sort out and tell/teach/train it which houses/images are for flipping and which aren't/are passes. but then upon examining the house images dataset we found, the issue then seemed to be that the images dataset was mostly of nicer homes, so they wouldn't be good when tryna distinguish b/w flip and not flip. the best way to train, at least initially is to make the line of demarcation very obvious…… or is it? maybe it's actually *better* to make it nuanced to make the machine more advanced right off the bat? hmmm

so i was getting ready to compile my *own* images database from MLSs of good and bad, but asked Raghu for advice and he suggested making the existing dataset work. and then actually when I looked even *closer* there were several hidden dumps in there! and also there's *wayyy* more images than we need - 2000 so he suggested narrowing it down to 200!

Another thing I started doing as I went thru the images to select/train/assign - so the dataset has 4 images for each house: frontal, bedroom, bathroom & kitchen. but in some cases one/some things are fine and not others. so idk if it'll mess it up but I'm thinking/realized - why make ourselves do whole sets? don't wanna confuse the machine - just pick the *best/worst* images overall to feed it - *cuz it doesn't matter* whether they belong to one house or many! it doesn't know - it's just looking for what to look for / what things to identify! so yeah, I started doing *partial* set matches

Noticed in a lot of these house sets, you really can't make a judgement from the outside alone. in fact, I think for the most part it's irrelevant. unless we're breaking up learning by facet/aspect and then combining/joining, there's not much need. I'm thinking I'm only gonna keep an exterior image if it's especially nice or especially bad. again, idk if it's gonna look at each house in/as a set of 4 or just a collection of interior images…. they probably did do the former, but I think I'll tweak it and do the latter. if I did my own custom set, I would've def included living room too

since there's so many images, I don't have to linger/study each one too long. in the/my first pass through I'm tryna move quickly and just jump on the *obvious* ones, in either direction

also there were a few instances of accidental duplicates, like where the file is labeled as 'bathroom' but it's just the same picture as the bedroom

also, have to verify, but almost positive I saw a few images more than once… >> *IS THERE A WAY TO CHECK DUPLICATE IMAGES???*

*>> did quick search:* [*https://www.google.com/search?q=IS+THERE+a+way+to+check+for+duplicate+images&oq=IS+THERE+a+way+to+check+for+duplicate+images&aqs=chrome..69i57j0i22i30j0i390l5.9466j0j7&sourceid=chrome&ie=UTF-8*](https://www.google.com/search?q=IS+THERE+a+way+to+check+for+duplicate+images&oq=IS+THERE+a+way+to+check+for+duplicate+images&aqs=chrome..69i57j0i22i30j0i390l5.9466j0j7&sourceid=chrome&ie=UTF-8)

*one result (address says 5 but title/actual is 13!):*[*https://wethegeek.com/5-duplicate-photo-finder-tools-to-delete-duplicate-photos/*](https://wethegeek.com/5-duplicate-photo-finder-tools-to-delete-duplicate-photos/)

your judgement to flip a place is often influenced by the décor/furnishing - like ugly furniture / too many colors going on, and messiness, AND *QUALITY OF THE PICTURE!* like many are cut off, blurry and/or *DARK!* so I wonder how the ML is gonna interpret those… cuz you'll never really see a very nice place with a *bad* photo. and similarly you never see a *nice,* updated place with bad décor. if you did get that rare, unusual, *made up* example to see if you could throw off the machine / stump it, you might be able to just from image quality alone. *#disguise*

and like are we gonna confuse it by having both stuff that is bright and dark, like how much weight will it assign to that,… and can we control that? like identify "brightness" and weigh it *down, well, lol,* actually the *opposite,* -> literally, bring its *weight* down!

to get advanced, do you have to like tell it to first identify the components - like these are the cabinets, this is the floor, which means that's the ceiling, which means these are the walls; this is the fridge, oven, etc. cuz like if gray wood floors are in, but a place has old school gray wood lined *WALLS*, you gotta be able to distinguish!

will it be able to distinguish between *old style* white cabinets vs new? maybe have to look at the knobs, well first identify the knobs as a separate, distinct part, to distinguish/differentiate them

so its interesting, within the broad/macro delineating flip v/ don't flip, you have to have many micro/*sub-*delineating - like a flow chart - like look at the cabinets - what color are they? brown? well now we gotta distinguish is it the *old* style brown or semi-modern style brown - later 90's/early 2000's. gotta pair w/ other factors. like biggest giveaway will be look at the *COUNTERS!* if they're *GRANITE* pattern, then *DEF* semi-modern! which could or could not be a flip, but for our purposes wouldn't be a flip cuz we're focused on the obvious and the biggest bang for our buck. to many ppl that would still be very nice, clean. but if *plain off-white* counters and like *small square-*tile floors and like bad/dim lighting and *OFF-WHITE FRIDGE to boot!, then add very heavy weight towards / mark flip/old!!!* are the cabinets white? okay then have to ask another question, cuz could go either way again. are knobs round or slender/slim/lined? round -> old, slender -> modern. also can factor in the lighting -> if it's bright too, it's much more likely / confident that it's modern. if it's round *AND* dim or grainy or generally labeled/flagged as bad picture quality, then it's very very likely that it's old/outdated. and also if there's brown wood trim present, or like tile floor or tile wall, or too many colors, it's most def outdated! if the fridge, oven (& microwave) are white, also can hint strongly at outdated. *IF THERE'S WALL PAPER!* if there are kitchen curtains, esp w//if they have patterns or are off/mulit-colored! can it distinguish b/w very old / 90s off-white counters for kitchen & bathroom vs. modern white-gray marble? again gotta support/buttress/fortify/bolster/supplement w/ help of other factors put together!

it's like, imagine if *ALL* you had was text, data based on the pictures, to go off of, but it was like objective and *YOU* had to make the judgement based off that. how well would you do? gotta think of it like that, cuz that's exactly what you're telling the computer to do / what you're giving the computer to work with for the final decision. *it'd be like:*

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Kitchen** | **cabinet color** | **cabinet knob shape** | **counter color** | **fridge color** | **dishwasher color** | **lighting** | **picture quality** | **neatness/**  **clutter** |
|  | white | long | white | black | black | sunny but dark | grainy | semi-messy |

*SEE HOW HARD IT IS TO IMAGINE THIS WITH JUST TEXT!!* you gotta keep looking at this / keep this in mind and ask yourself: *is this enough to go off of?* **could a *HUMAN* make a decision based off of this??** cuz if a *HUMAN* can't then a machine *DEFINITELY can't!!!* it's only as smart as you program it to be!!!

Here's what I was *TRYING* to describe / capture above! - below:

A kitchen with a black refrigerator

Description automatically generated with medium confidence

lol it's funny how there's this very real trend of "sunny (as seen/can be telled from view/glimpse of little window(s) but *still* dark!" *>> red-flag!*

can identify a bad picture / too zoomed in by seeing how many elements it identifies!! and *which* elements! make sure it captures all the cabinets, the fridge, oven, sink, etc. like determine which ones it *MUST* have in order to be considered complete!

aH I can see this being very cool! like just like I imagined - tryna replicate the human brain to a computer / map / translate / osmose / transfer / *teach* a computer/machine to *learn!*

make it distinguish between *ARTIFICIAL* / interior lighting and ceiling light! like can train it to look for the ceiling, then look for *lights* on the ceiling, i.e. *bright spots,* and then it'll know. vs. if it's bright *WITHOUT* bright spots on the ceiling!

I'm seeing examples of very nice looking homes on the outside, but w/ some very ugly elements on the inside

*IS THE MACHINE ABLE TO HANDLE DIFFERENT TYPES OF MODERN???* like can it classify/case-when? like there's the gray modern, the dark brown modern, the honey brown modern, the light/sawdust brown/all-white modern, farmhouse modern, brass/marble modern, etc. cuz if it tries to unify *ALL* of them on/to one common ground *it's gonna get very confused!!!* does it *automatically* break into (sub-)categories/*CLUSTERS* like we saw w/ that one analysis *(PCA??)* in Big Mountain!!!)?? like:

A white paper with writing on it

Description automatically generated with low confidence

(note: I added more to this markup on my physical / original paper since)

training/teaching a computer is a very interesting/profound human experience. because as humans one of the beautiful things Allah blessed us with is intuition, which we can often rely on. which is basically shortcuts where our brains take care of complex algorithms without us having to think about it much. like how is it that when we look at someone, we can instantly identify (for the most part), their race, age, etc? like do we consciously have an internal (or external lol) dialogue (well monologue i guess lol) with ourselves like, okay well I see wrinkles, gray hair, paired w/ old ppl style clothing… like our brain does all that for us nearly instantly. it's cuz we've seen so many examples, we've trained ourselves ever since we were kids. we were first told who was who, like "this is grandpa," this is "uncle," etc - ppl who we then learned were much older. we imputed that they're much older, and so this is what "older" people look like. we have all that in our brain so that now, w/ our *test* set, we can very accurately identify who's old and who's not. there will always be the exceptions of course of ppl w/ maybe special conditions or *very good genes /* very healthy who don't show the traditional signs of age that stump us. but so when we rely on our intuition for these autonomous / involuntary decisions/analysis/conclusions, we don't often think of *why* we've made the determinations we do. so the very cool thing w/ machine learning is that it *forces* you to consciously step back and think, what *is it* that actually makes this / makes me think/know that this house is old and needs renovation? and also it's good cuz it forces you to separate out the things that are less objective and more emotional / maybe extraneous, like unagreeable décor/style/furnishing/picture quality - these things that make you angry/disgusted, get you emotional, and make you exaggerate your judgement, *"like seriously you couldn't take a few minutes to tidy up before taking the picture?!"* or you couldn't double check how the picture turned out / are you telling me you looked at this picture and thought this was presentable to post for the world to see / thought you were going to get a good offer w/ these pictures?! couldn't wait a few hours till the morning to take a nice bright picture during the day?? how did this look okay/attractive to you! seriously?!!-you can see yourself in the mirror w/ your cell phone and feet!! *eww it looks like a grandma lives here and has had the same stuff for 60 years!!* but yeah by having to manually study the patterns, that helps you *explain WHY* you make the judgements you do so you can (1) justify it and (2) teach/explain it to somebody else, or in this case, some*THING* else, *i.e. THE COMPUTER!!!* #machinelearning is paired w/ /necessitates #humanteaching / instruction

see whether it's OPEN CONCEPT OR NOT!! island/island size! cabinet size

is there a backsplash??

I see some good examples but they'd be confusing cuz have a mix of elements going on. like more advanced, nuanced, special situation like where it'd be like - *but I thought you told me to reject funky colors!/that funky colors are BAD! >>* You: *yes I know, usually! but this is an ACCENT color!* that'll be too hard to train for now

I guess the emphasis should be on what to *flip*, more so than what *not* to flip - cuz the target is what to flip. and it's like - everything *ELSE* is 'not flip'. also, the modern/not flip stuff tends to be more consistent w/ the styles, easier to identify, so maybe needs less training/examples? whereas there's *LOTS* of examples of what needs flipping - cuz there's *MANY* ways to do something *WRONG* - but only a *FEW* to do them *RIGHT!!! #deep* so we need to feed many examples of what to flip to make it clearly understand

Can I mass rename stuff? and also filter? cuz like it'd be cool to remove the frontal pics from consideration. and be able to sort so that I could look at *all* the kitchen pics at once. also the order within each group is annoying, like since I don't want/'m not as interested in the exterior pics, it's annoying that they come inbetween the bed/bath and kitchen pics! interrupts! is what I wanna do doable via python?

if something is older style but in good condition, clean and good pic, if it's bright, doesn't seem so bad

oh actually, I realize now that sometimes you *do* see modern updated stuff that has bad pictures that really undermine/undersell it. could be new construction or flips where the person is just lazy, or it's the contractor / contstructor himself taking pics off his phone at night who has no idea how to take a good pic, thinks the work will speak for itself, maybe cuz he doesn't wanna go thru an agent so doesn't have to pay commission

you gotta be careful what you feed the model cuz you don't want it to get distracted by / latch on to miscellaneous / extraneous things that are besides the point / not the focus

interesting, when I first looked at this dataset / esp the beginning images, I was like these houses are all *too* nice/updated… but actually, now that I've gone thru it, *THERE'S HARDLY ANY TRULY MODERN ONES!!!!* it's these weird California homes where there actually might not even be a *SINGLE* "standard" modern home that you'd normally see most places, like w/ the gray/light blue; or crystal/bling, or brass/marble, etc. what we actually have, esp/atleast in the beginning, is a lot of *SEMI-MODERNS!* so that's a problem cuz there's not nearly as much a difference between outdated and semi-modern and the point w/ machine learning, at least at this stage, is to *make it very obvious / clear for it!!!* and the other thing is that, in real life, semi-modern is a toss-up -> there's a range but it's very hard to tell from pictures *EVEN FOR A HUMAN!!!* you really have to see it in person to decide whether this would be worth flipping or whether there'd be a good number of ppl who'd be happy with this as is. there's diff kinda flips - there's like rags to riches complete rehabs, and then there's "updating," from like last decade-/20-30 years ago/last-gen to now. like from complete gutting and *replacing* everything to simply repainting more neutral colors

man, now I'm honestly thinking maybe I would be better off getting my own pics…

look for databases of modern pictures online. *OR -* don't even have to look for a *SINGULAR* databse, literally just google search and save pics from there! like fixer-uppers, rehabs, modern homes, and actually, *THE BEST MIGHT BE BEFORE AND AFTERS!!!! THAT WAY* YOU'RE COMPARING LIKE FOR LIKE SO YOU CAN SEE *EXACTLY* what the key differences are to focus on!!!

so bc of the above^ it's taking me some time to sort through this set cuz no clear line. so I'm thinking, might as well get good pics from online. but that might take me some time to get the amount I need. but actually, if I get custom curated pics, I can probably get by w/ much less, or at least some less. cuz I might as well do this the way I want that's practically useful for me the first time!

can get, say, 10 bedroom pics, 10 bathroom, 10 kitchen, 10 living, 10 exterior - each for training & for test. like said earlier, don't need as many examples of nice ones. and I think we have plenty of pi

I thought I'd find much more on google image search, but for some reason not what I was looking for. not like real estate type pics. so best bet is I may need to go straight to mls's

so right now the way it's shaping up is that I have the training as almost all semi-moderns, but a bunch of true moderns in the testing, so will be interesting to see!

I think I figured out the trick - when looking at these pictures, gotta constantly ask yourself *ONE* thing / the same thing every time - would I flip this? that goes a long way in helping you decide w/ all these mid-moderns. cuz that's what determines whether something is gonna be worth the time / return. cuz think about it - if something is semi-updated/modern, *SOMEONE* will prob still want it to live in, will be desirable enough, so price will be semi high. cuz think about it - it's not just about *can* you update it. cuz even w/ stuff that *IS* updated, *it may not be your style and you wanna change it all even tho / if it's brand new!* and obv it's gonna be so expensive so doesn't make sense to get it if you know you wanna change it all! like if it's already in good shape why would you break it all down and scrap it! so if you're gonna replace you might as well do something that's broken! so you gotta think, if you're gonna spend the time and money to *REPLACE THINGS,* like get the new cabinets, counters, floors, lights, etc - all that money for labor and parts, like if you're gonna *REPLACE* stuff anyway, *you want it to be the oldest, trashiest, ugliest stuf possible!!* to get the most bang for your buck! cuz that way no one's gonna want it, and you'll get it for cheap, so the replaced stuff will give you a much bigger profit markup/margin! and also esp if the pics are bad no one's gonna want it/ even wanna see it! now it's also true that for houses that aren't in *as* bad shape or *as* outdated, they may be bigger and/or because they're in better shape the *cost* also won't be as much to fix, and also they may likely be in a better neighborhood too!!! and as far as the parts you're replacing - maybe it's possible to resell them for something? not sure how that works. remem - location location location. so in that sense, it could still make sense to get something that's not necessarily a fixer-upper foreclosure! cuz think about it - it wouldn't make sense to get a run down foreclosure in the hood and make it an ultra-modern ultra-luxurious bungalow! cuz that's a huge mismatch!! the type of ppl who would want that style would never want that location! gotta look at schools if it's targeting families! so actually in that case then renovating a semi-updated, if the neighborhood is desirable, could actually make a lotta sense! but that would be for a separate algorithm that looks at text and factors in school ratings and crime rates

so, for our purposes here, we're looking for stuff that looks *bad.* cuz like we mentioned above, there's 2 kinds of flips, so we're gonna focus on this kind, of clearly bad stuff

also remem - you don't have to get it perfect the first time! just try one way and see how it goes and then tweak if you need!

it's easy for you to get confused w/ a project like this cuz your mind flip flops around to diff places - sometimes you're thinking about what you PERSONALLY like - again, it's like we said earlier, your emotions come in the way, sometimes you flip flop on what kind of houses you're tryna flip, your standard/line fluctuates, could be depending on your mood, energy, hunger, tired of staring at a screen/brain fried! lol. are you training the machine to *think like YOU -* like a recommender engine, driving style adapter? or do you have to think in the shoes/mind of the *INVESTOR,* the target end-user?

ask - would someone want that, as is, to live in? would a good chunk of ppl think this is nice? then leave it

So now I need to figure out what to do. I could spend time trying to figure out how to make my model perform better, like looking at all those possibilities I laid out. we can go over that tomorrow by looking at what I might be doing wrong, what could improve, compare to ahmed/moustafa's