

Capstone Project Proposal



Shasu Vathanan

Business Goals

Project Overview and Goal

What is the industry problem you are trying to solve? Why use ML/AI in solving this task? Be as specific as you can when describing how ML/AI can provide value. For example, if you're labeling images, how will this help the business?

To detect Face Mask in public gatherings. Using a streaming video camera on the street.

Using Artificial Intelligence, people group has additionally been a piece of these undertakings. Specifically, improvements for checking social distancing or distinguishing face masks.

With an end goal to paint a more complete picture, I chose to show the innovative procedure behind an answer for a simple use case in computer vision:

Goal:

- To Identify the public face and their face mask usage.
- Identify people that pass through the streaming video cameras on the street.
- Collect the data, % of people wearing masks and not wearing a mask.

Business Case

Why is this an important problem to solve? Make a case for building this product in terms of its impact on recurring revenue, market share, customer happiness and/or other drivers of business success.



In spite of the fact that the difficult explanation is comparative (recognizing face mask use), the truth of images acquired from certifiable sources like CCTV (Closed-circuit TV) or reconnaissance cameras can be a lot harsher. During the greater part of the turn of events, however we need to test it with different sources, similar to the streams appeared toward the start of the post.

- Live face extraction and coordinating to

| | |
|--|---|
| | <p>databases of thousands of countenances inside 3 seconds.</p> <ul style="list-style-type: none"> • Handling various streams and different countenances per image. • Face match rate > 90% with great camera edges and lighting. <p>Are a SaaS stage and the client are charged based on the quantity of cameras and use-cases running for them? The essential cost is around Rs 1,000 for each camera for each month yet can go up to Rs 3,000 for each camera for each month.</p> <p>Bootstrapped to date, Integration Wizards professes to be gainful from the principal year of tasks, having an ARR of \$3.2 million. Proceeding, our startup intends to expand its utilization case and client traction and manufacture a worldwide accomplice arrange. Our startup is likewise wanting to band together with different establishments to enable de-to heighten the danger of contamination openly with the lockdown being lifted in different pieces of the world.</p> |
| <p>Application of ML/AI</p> <p>What precise task will you use ML/AI to accomplish? What business outcome or objective will you achieve?</p> | <p>The framework works by performing two primary undertakings:</p> <ul style="list-style-type: none"> • Object detection with a neural network (SSD), pre-trained for face detection. Output: list of bounding boxes around each detected face. • Classification in two classes (with/without mask), using another neural net. Output: score from 0 to 1 signifying the probability of a face wearing a mask. <div data-bbox="829 1266 1258 1566" data-label="Image"> <p>PLEASE WEAR A MASK</p> </div> <p>The pre-prepared face recognition model appears to work extraordinary for his case, and it identifies faces in any event, when they are incompletely secured by covers. Thus, no compelling reason to re-train anything for the principal model like object identifier, SSD (single-shot detector)</p> <p>This peruser you had always wanted associate made</p> |

| | |
|--|---|
| | this dataset with almost ~1400 pictures in an extremely sharp and successful manner. We have to use dataset with customary appearances, and misleadingly added veils to a portion of the pictures. Then we utilized other PC vision strategies to naturally put the covers over the countenances. |
|--|---|

Success Metrics

| | |
|--|---|
| Success Metrics What business metrics will you apply to determine the success of your product? Good metrics are clearly defined and easily measurable. Specify how you will establish a baseline value to provide a point of comparison. | <p>How tracking works:</p> <p>The tracking algorithm lives during the entire video arrangement and is refreshed bit by bit, with each new casing that is handled.</p> <p>At each progression, it gets a rundown of focuses and attempts to coordinate each point with recently identified focuses. For that, it can appraise the new position that each point ought to have, as indicated by its assessed speed.</p> <p>At the point when it identifies that specific focuses (e.g: relating to an individual's eyes) appear to have "made due" for a few stages, you can utilize it to state: this is a solitary individual and it's moving from here to there.</p> <p>Note: The tracker we will utilize couldn't care less about what's really on the picture. It just gets a rundown of point facilitates at each progression, and predicts their future positions. There are no PC vision errands here; those are just acted in different parts.</p> <p>Essentially, we have to change over the posture data into a crate, which can be characterized with two boundaries:</p> <p>Box size:</p> <p>Utilizing head keypoints: in view of the separation between eyes, ears, and nose.</p> <p>Utilizing chest keypoints: in view of hip-to-neck separation appears to be more steady for individuals glancing in various areas.</p> <p>Box position:</p> <p>Utilizing head keypoints: the reference is a point in the focal point of the head recognized focuses.</p> |
|--|---|

Utilizing chest keypoints: simply utilize the neck position as a source of perspective.

Data

Data Acquisition

Where will you source your data from? What is the cost to acquire these data? Are there any personally identifying information (PII) or data sensitivity issues you will need to overcome? Will data become available on an ongoing basis, or will you acquire a large batch of data that will need to be refreshed?

In any case, so far, I just changed the item identification module: rather than utilizing a SSD neural system to identify heads, we're utilizing present estimation and removing heads from that point.

In the wake of acquiring the bouncing boxes for the heads, the methodology is essentially the equivalent: we have to do mask/no mask grouping on the head pictures.

Face detector:

I had video transfers and a content that identifies faces on them. It was direct to extricate these faces naturally and spare them as individual pictures. At that point the framework simply expected to physically place them in two envelopes with mask, without mask. Furthermore, there was a third envelope: unknown, it was gigantic yet had a short life.

Mask/No mask dataset:

Even though we still everything has a similar elevated level parts, the idea of the image is totally extraordinary and the falsely produced dataset would not work. As I said previously, faces are never frontal to the camera, image are the much lower size and quality, faces are foggy and covers as well, and they can come in numerous potential surfaces and hues.

Data Source

Consider the size and source of your data; what biases are built into the data and how might the data be improved?

Data Source:

- Current state of the art (SOTA) models for image classification on CIFAR.
- Google: Big Transfer (BiT).

Specifically, we could utilize the BiT-M model for which there are delivered loads, pre-prepared in the dataset ImageNet-21k which contains 14 million pictures from ~21k various classes.

Choice of Data Labels

What labels did you decide to add to your data? And why did you decide on these labels

Data Label:

Without Mask: Identify the people without the mask.

With Mask: Identify the people who were a mask.

versus any other option?

Other: To identify others (like animals, helmet, scarf etc.)

Model

Model Building

How will you resource building the model that you need? Will you outsource model training and/or hosting to an external platform, or will you build the model using an in-house team, and why?

- From the outset, it is anything but a conspicuous necessity, however utilizing following is really a distinct advantage. We should recap why:
- The classifier can be run a few times on a similar face, and we can take the most elevated scores to settle on a choice about every individual; i.e: watch the whole video arrangement like we would do as people, rather than individual disengaged outlines.
- We can have measurements dependent on the real number of individuals that cruise by, freely of the time that they show up on the screen. Rather than taking occasional screen captures and afterward taking some normal.
- Quicker handling, since we can abstain from processing the posture estimations in each casing, by inserting middle of the road positions.
- Higher heartiness to awful identifications that could show up at certain casings.
- Smoother face boxes

Evaluating Results

Which model performance metrics are appropriate to measure the success of your model? What level of performance is required?

A basic standard I use is to characterize just when there are at any rate three facial keypoints (to pick between the two ears, the two eyes, and nose) over some edge. That is the means by which we abstain from running the classifier for heads without an obvious appearance.

In the accompanying video, we're utilizing this technique. Over each head we show the most recent yield probabilities of the classifier:

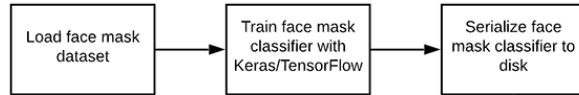
- Close to 100 methods a cover was distinguished.
- Watches out for 0 if no veil is distinguished.
- Around 50 if uncertain

Minimum Viable Product (MVP)

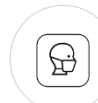
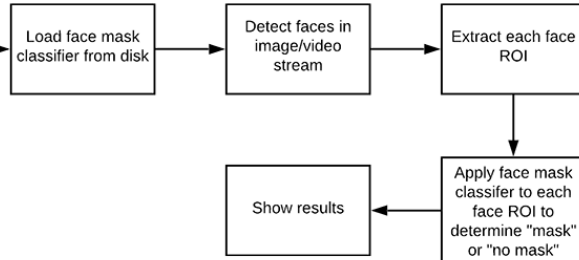
Design

What does your minimum viable product look like? Include sketches of your product.

Phase #1 :Train Face Mask Detector



Phase #2: Apply Face Mask Detector



Subscribe to Face Mask Alert



Add your existing Camera

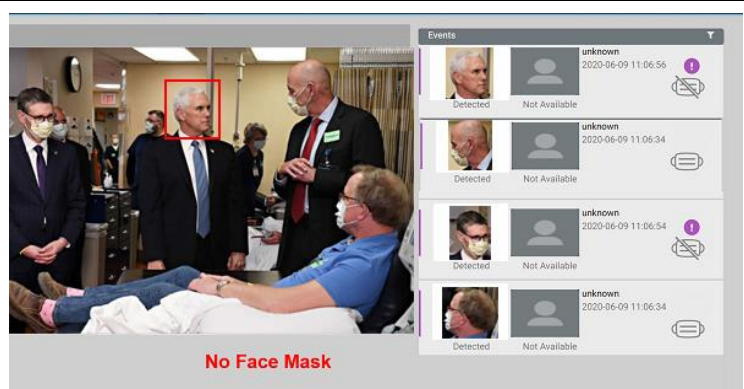


Start getting No-mask data and alerts



Send sms alerts to Adminia and Violators

Face Mask Detection System utilizes existing IP cameras and CCTV cameras joined with Computer Vision to recognize individuals Face with or without Mask.



Face Mask Detection Platform utilizes Artificial Network to perceive if a client isn't wearing a mask. The application can be associated with any current or new IP mask identification cameras to recognize individuals without a cover. On the off chance that the camera catches an unrecognized face, a warning can be conveyed to the executive.

Use Cases

What persona are you designing for? Can you describe the major epic-level use cases your product addresses? How will users access this product?

Tested several methods and the best results I got were simply using a thresholding + voting approach:

Thresholding: only consider classifier outputs with a score under 20 (no mask) or above 80 (mask). Discard the others.

For each person, count the number of “votes” for the mask (positive, green) minus no mask (negative, red) as they walk their path.

After at least 3 votes, decide if it's a masked person based on the balance of the vote.

Limit votes to 100 for the faster reactions if the person puts on or removes the mask during the sequence.

Avoid side faces:

In spite of being valuable, the classifier scores aren't altogether solid. Watch the resting fellow 2: he doesn't turn his head totally and even so the classifier begins working.

To sift through these edge cases, we wound up requesting the two eyes to be obvious, which has the impediment that side appearances are doubtlessly disposed of, regardless of whether now and again the cover is discernable. We in reality left this as a configurable alternative, however it appears to work better empowered. Recollect that we despite everything have the entire grouping for every individual, so we can be critical here.

Roll-out

How will this be adopted? What does the go-to-market plan look like?

Market analysis:
Customer needs, Market dynamic, Company capabilities, Competition, Collaborators, and partners.

Market selection
Market segmentation, Market selection, Segment targeting, Product positioning.

Marketing Mix:
Product feature, Price, Advertising, Process.

Customer:
Awareness, Interest, Evaluation, Commitment, Referral, and Customer loyalty.

Post-MVP-Deployment

Designing for Longevity

How might you improve your product in the long-term? How might real-world data be different from the training data? How will your product learn from new data? How might you employ A/B testing to improve your product?

Trying other models for pose estimation:
In this execution, we utilized Open Pose in light of the effortlessness (and the way that we had utilized it previously). However, this is an open territory of examination, and there are some new improvements that clearly work lighting quick, even in CPU. We should try different things with a few executions to perceive how they charge.

Improve the classifier:
We have seen others really utilize 3 classes for order: with mask/without mask/misplaced mask. We figure this methodology would bode well since our classifier experiences some difficulty with these cases: the cover is distinguished yet additionally the nose and mouth are obvious, for instance. We didn't add such cases to the dataset on the grounds that we didn't know it was a smart thought to place them in a portion of our 2 organizers, so this new methodology would comprehend the issue conveniently. We can likewise source more pictures from more video transfers and assemble a superior dataset.

Ditch pose and try out robust head detection:
Another fascinating methodology is to utilize some single-shot article identification strategy to recognize any sort of head, in any conceivable point, with names like "front face, with mask", "front face, without a mask", and "face not obvious". We have to follow heads rather

| | |
|---|--|
| | <p>than just faces, to have a ceaseless development regardless of whether the individual looks the other way than the camera. This general methodology probably won't be as powerful as the current one, yet would most likely be less complex and quicker, and it may be a smart thought for gadgets with restricted assets.</p> |
| <p>Monitor Bias</p> <p>How do you plan to monitor or mitigate unwanted bias in your model?</p> | <p>Re-identification methods can make the framework stronger to interferences and impediments, by separating picture highlights from each followed individual, to permit coordinating individuals regardless of whether the screen freezes and they re-show up somewhere else. Recall that the current methodology expects individuals to move constantly. Nonetheless, this method includes a huge layer of unpredictability and computational weight.</p> <p>As should be obvious, there is no silver projectile here, and the methods will shift as indicated by the prerequisites of the arrangement and the limitations, similar to the earth in which it should run. Vigorous video examination is a long way from illuminated, however steady advances can be taken to improve any piece of the arrangement.</p> <p>We want to keep taking a shot at this, and expectation you had some good times while perusing our post! If it's not too much trouble stay tuned and don't stop for a second to post your remarks, questions or proposals!</p> |