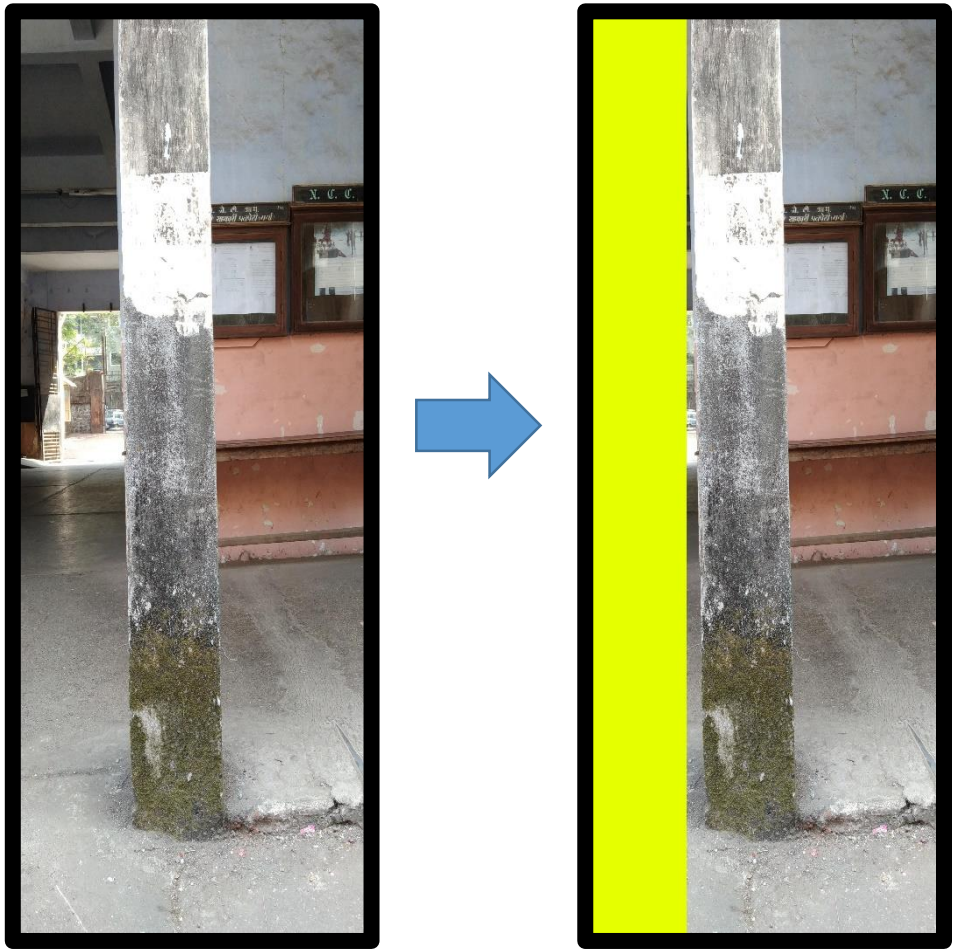
Dataset Preprocessing Steps

Original Image

- Let us take an example of the given image

Ask user for the crop window

- It will start the cropping window from the topmost and leftmost unvisited sub image
- This cropping window is to ensure that the sub image is pure (not a wall, wall with no defects, damp wall, cracked wall, paint peel off)
- Adjacent image shows the first cropping window which is purely “not a wall” class



Ask the user to tag the pure sub image

- This tagging, although being manual would be done only once for a bigger sub image rather than many smaller sub images
- Adjacent image shows that the user has tagged the sub image to the “not a wall” class

Process the sub images

- The images in the tagged sub image would be cropped into all combinations of square images starting from square of size 256 and incrementing the square size by 1
- This process would be done by background threads and a thread would proceed with the next step
- If the user entered the class of a “not a wall”, then no processing would take place and the complete sub image would be dumped

Ask user for the crop window

- It will start the cropping window from the topmost and leftmost unvisited sub image
- This sub image again is pure with only the “wall with no defects” class

Ask the user to tag the pure sub image

- Adjacent image shows that the user has tagged the sub image as “wall with no defects” class

Process the sub images

- The crop window that was selected in the previous step was a “not a wall” class
- The images in the tagged sub image would be cropped into all combinations of square images starting from square of size 256 and incrementing the square size by 1
- This will again be done in a background thread
- Only one manual input is required to obtain these all sub images

And so on…

- Likewise the processing would continue until all the sub images of the image would be tagged
- These steps help us to reduce the manual entry part as less as possible
- These steps also give pure dataset with almost no noise that would help increase the machine learning model to train with higher accuracy