Data Augmentations

For data augmentation, I think there are a couple things that we should consider. I list them here and my reasoning is below:

1. Image Resizing
2. Rotation
3. Shearing
4. Brightness
5. Noise
6. Exposure

**Image Resizing**: We need a standard image size, so this normalization step is absolutely necessary no matter what augmentations we perform next.

**Rotation**: Not every image we receive will be from the front looking at the foods, or perfectly from the top down. Thus, I believe we should add rotation to account for the different angles that we may receive images from.

**Shearing:** Similar to rotation, not every image will be perfectly from the top down. By applying shearing, we can account for images that may be taken from a 45-degree or some other angle, and thus will make our model more flexible when it comes to real user data from the DRC.

**Brightness**: Although we emphasized lighting for the DRC partners when we discussed image collection guidelines, I still believe that there will be differences in brightness, especially when we have images from the women who may be less experienced with using mobile devices and taking pictures. I think we are likely to see some images that are too dark or too bright, and thus we should account for a difference in brightness. It doesn’t have to be drastic, but a 10% difference in both ways should be enough to account for this difference.

**Noise**: Roboflow adds noise by introducing a graininess to the images. Since we cannot be sure of the resolution of images taken by our partners in the DRC, I believe that adding noise is important. We cannot expect every image to be 1080p, and our model should still work on images that may present in 360p or 480p, and the way we can do this is by adding pixelation/noise through this augmentation step.

**Exposure:** Similar to the brightness above, some pictures taken from the DRC may be over or under exposed. While I don’t think this will be a huge issue since we have mobile devices that should be able to adjust for exposure automatically, I still think we should account for a little bit of exposure difference in both ways to make our model more resistant.