Soil erosion detection solution report

Solution description

1. Discard all parts of the erosion mask that are outside the territory depicted on the aerial image.
2. Divide the aerial image into 128x128 3-channel images (features) and divide erosion mask into 128x128 boolean matrices (labels) 7255 samples.
3. The model architecture used is U-Net with encoder (contracting part) weights initialized using MobileNetV2. The model has 5 convolutional layers in the contracting part and 5 in the expansive part. The output of the model is a 128x128 matrix of logits (values from range [-inf; +inf]).
4. Perform transfer learning on U-Net with encoder weights freezed. The loss function used is crossentropy.

My proposals

1. The aerial image that was given to me has 3 channels corresponding to 3 response curves. Increasing the number of channels (bands) in the aerial image might improve the detection accuracy. I think that it is important for these response curves to be independent (no linear combination of response curves exists that is equal to 0 response curve), so that each band gives new information.
   For example, images used for land use classification in [1] consist of 4 bands (RGB + panchromatic).
2. Different aerial images might have been shot with different meters/pixel value and different number of bands and/or response curves.
   Therefore, in order to reduce the complexity of the mapping to learn, we might want to restrict the training, validation and test sets to the images of specific meters/pixel value and response curves. Thus the different datasets should be properly integrated by upscaling/downscaling and their colorspace corrected if it is possible.
   The second option might be to train on the distribution of image sets with various meters/pixel values and colorspaces. This might be achieved via integrating multiple datasets of aerial images without transforming to common reference. Also this might be achieved by performing augmentation: changing colorspace and upscaling/downscaling. This approach might eliminate the need for strict meters/pixel value and color space requirements when utilizing the trained model.
3. We could decompose the problem into the following parts:
   a. segmentation of the aerial image (field/nonfield or different land use types)
   b. soil erosion detection applied only to the parts of aerial image that is classified as field

   This might simplify model training: instead of learning one complex mapping, we make two models learn simpler mappings.
4. We could utilize other architectures. e.g. [1] extends DeepLabv3 architecture for performing multiclass segmentation of aerial image and achieves 70% pixel accuracy on images not involved in training.
5. [2] utilized also utilizes U-Net, however images are of higher spatial resolution (0.5-2 m compared to our 10 m) and the features include additional 3 planes derived from

DTM information. This suggests that we could also use other information beside RGB spectral bands.

References

1. https://www.researchgate.net/publication/360193574_Application_of_Deep_Learning_in_Land_Use_Classification_for_Soil_Erosion_Using_Remote_Sensing
2. https://www.mdpi.com/2072-4292/12/24/4149