Can you see the (linguistic) difference?
Exploring mass/count distinction in vision

Student: D. Addison Smith

Supervisor: Raffaella Bernardi, Co-supervisor: Iñaki Inza, UPV/EHU

Does the linguistic distinction between mass nouns (e.g. coffee) and count nouns (e.g. bicycle) have a counterpart in vision? This thesis explores the countability of nouns in language from the visual perspective. A dataset is constructed from images labeled for their English countability, with cropped objects denoted as either mass or count. Corresponding visual features are extracted from a deep Convolutional Neural Network (CNN). Since *count* nouns typically refer to well-defined, quantifiable objects, with *mass* nouns prototypically including less countable substances, the extent to which the linguistic distinction is grounded in the corresponding visual representations is explored. The hypothesis is that images of prototypical mass objects exhibit lower visual variance than do count objects. To test this we run two experiments which show that this linguistic distinction indeed has a counterpart in vision, mass nouns being more homogeneous in nature and externally more consistent across discrete instances (lower overall variance), with count nouns comparatively more heterogeneous, comprised of distinct parts, and varying more between individuals (higher variance). Variance is computed across various CNN layers and is indicative of the categorization when low-level features of the images are used, whereas any effect disappears when experimenting with higher-level, more abstract representations. A third experiment trains a neural network to learn and demonstrate a characteristic feature of mass nouns which is that they are arbitrarily divisible, i.e. a mass item such as *flour* can be split into arbitrarily many pieces and still maintain its "flourness," while this property does not hold for count items such as *bicycle*.