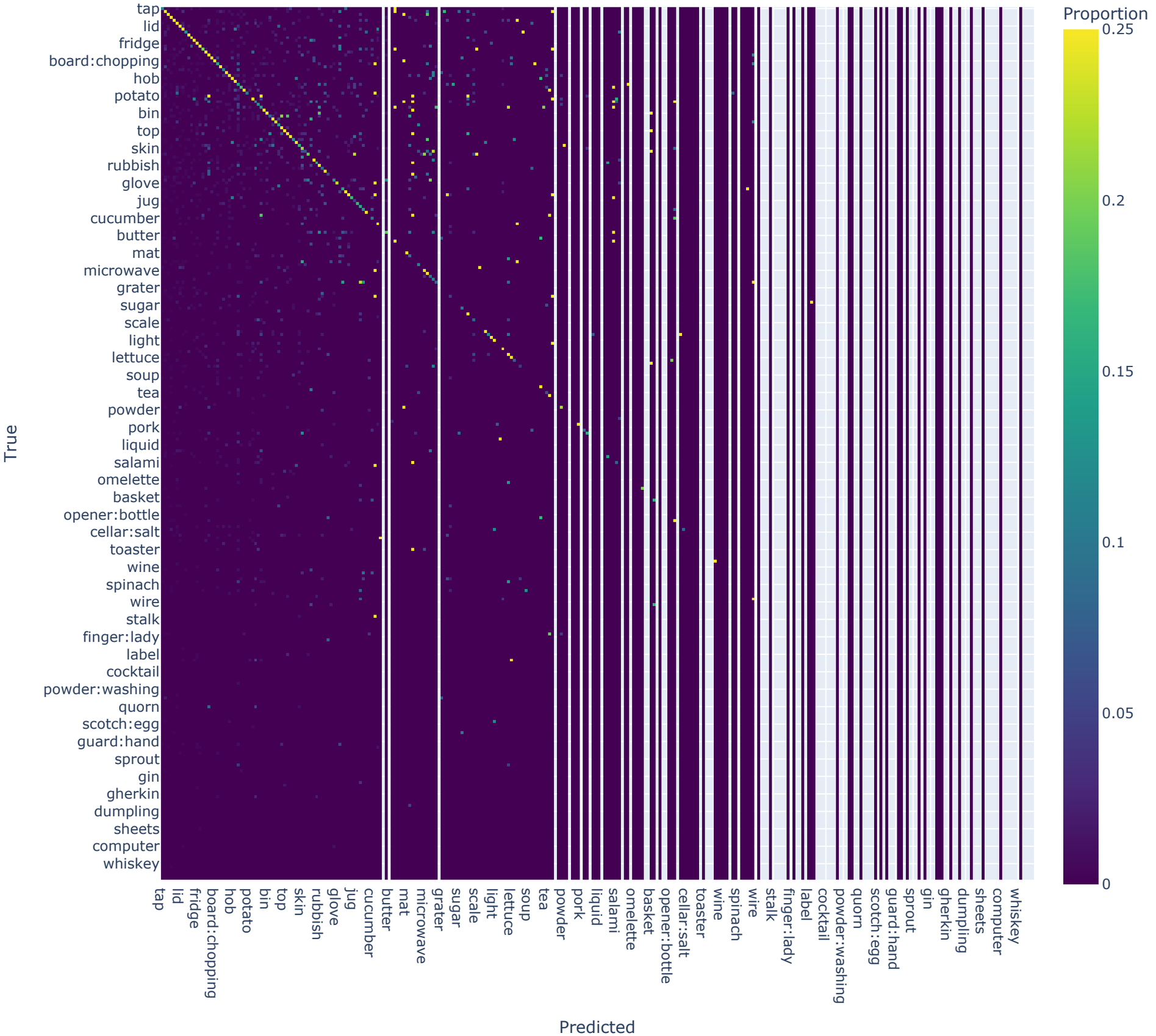Difference in confusion matrix for the top None noun classes between **multi-modal ensemble** and AudioSlowFastGRU