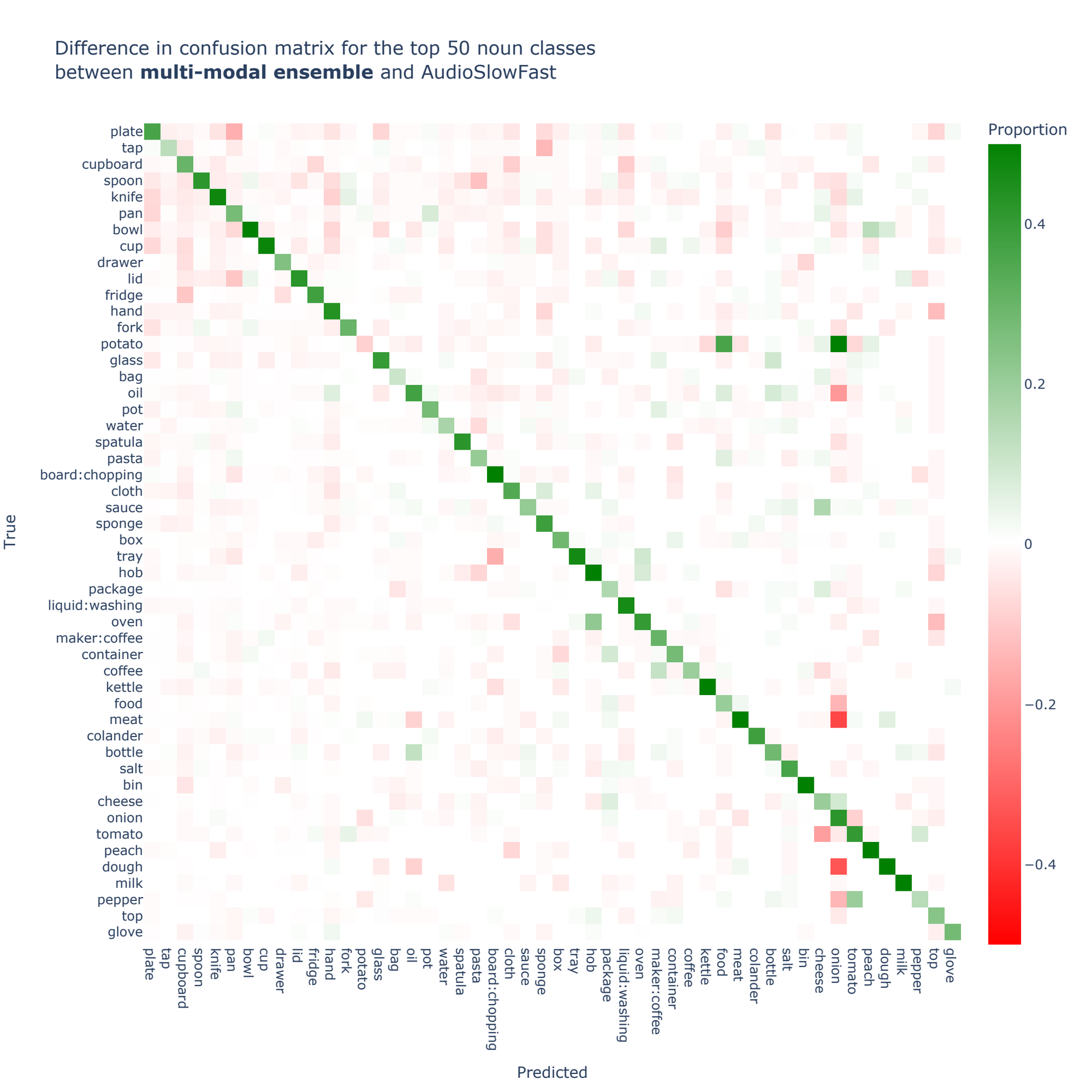
Difference in confusion matrix for the top 50 noun classes
between **multi-modal ensemble** and AudioSlowFast
True
Predicted
Proportion
0.4
0.2
0
−0.2
−0.4
plate
tap
cupboard
spoon
knife
pan
bowl
cup
drawer
lid
fridge
hand
fork
potato
glass
bag
oil
pot
water
spatula
pasta
board:chopping
cloth
sauce
sponge
box
tray
hob
package
liquid:washing
oven
maker:coffee
container
coffee
kettle
food
meat
colander
bottle
salt
bin
cheese
onion
tomato
peach
dough
milk
pepper
top
glove