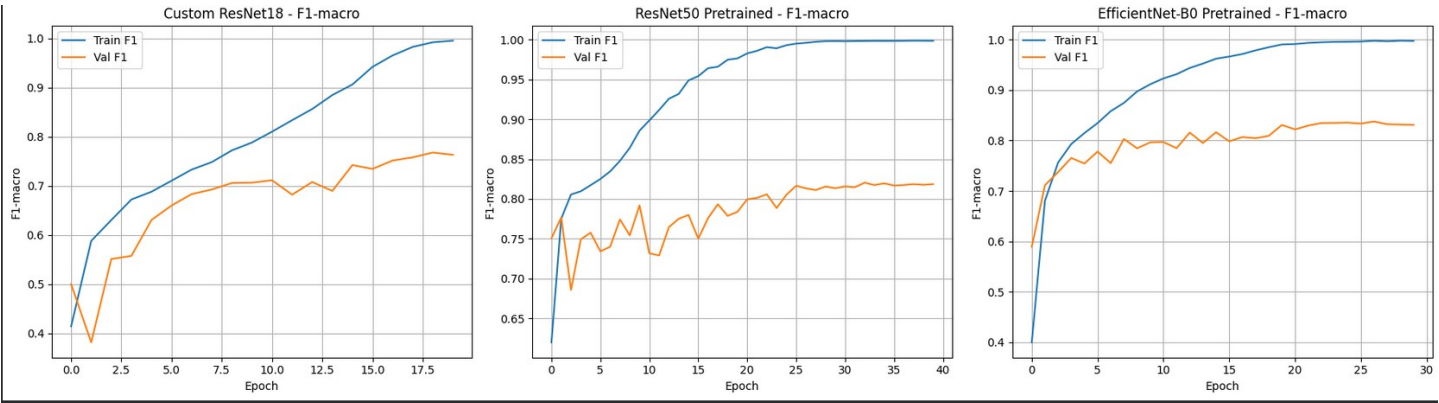


После некоторых правок (основное — снижение количества классов до 15, классы, имеющие до 500 экземпляров не учитываются) и повторного обучения результаты на вал.выборке стали лучше:



Evaluating Custom ResNet18 (from scratch): 100%|██████████| 6/6 [00:07<00:00, 1.28s/it]

Results for Custom ResNet18 (from scratch)

F1-macro: 0.7678  
F1-weighted: 0.8667

Classification Report:

	precision	recall	f1-score	support
Beige	0.59	0.53	0.56	30
Black	0.89	0.93	0.91	716
Blue	0.90	0.87	0.88	424
Brown	0.55	0.53	0.54	45
Green	0.90	0.72	0.80	39
Grey	0.82	0.84	0.83	474
Orange	0.77	0.71	0.74	28
Red	0.95	0.96	0.95	305
Silver	0.83	0.87	0.85	389
Unlisted	0.40	0.18	0.25	76
White	0.93	0.94	0.94	470
Yellow	0.94	0.97	0.96	33
accuracy			0.87	3029
macro avg	0.79	0.76	0.77	3029
weighted avg	0.86	0.87	0.87	3029

Evaluating ResNet50 (pretrained): 100%|██████████| 6/6 [00:08<00:00, 1.44s/it]

Results for ResNet50 (pretrained)

F1-macro: 0.8204  
F1-weighted: 0.9027

Classification Report:

	precision	recall	f1-score	support
Beige	0.73	0.73	0.73	30
Black	0.92	0.96	0.94	716

Blue	0.92	0.91	0.91	424
Brown	0.73	0.67	0.70	45
Green	0.88	0.77	0.82	39
Grey	0.86	0.87	0.86	474
Orange	0.82	0.82	0.82	28
Red	0.96	0.97	0.97	305
Silver	0.90	0.91	0.91	389
Unlisted	0.50	0.20	0.28	76
White	0.96	0.98	0.97	470
Yellow	0.89	0.97	0.93	33
accuracy			0.91	3029
macro avg	0.84	0.81	0.82	3029
weighted avg	0.90	0.91	0.90	3029

Evaluating EfficientNet-B0 (pretrained): 100%|██████████| 6/6 [00:07<00:00, 1.30s/it]

=====

Results for EfficientNet-B0 (pretrained)

=====

F1-macro: 0.8377  
F1-weighted: 0.9041

Classification Report:

	precision	recall	f1-score	support
Beige	0.86	0.83	0.85	30
Black	0.91	0.97	0.94	716
Blue	0.92	0.90	0.91	424
Brown	0.71	0.76	0.73	45
Green	0.97	0.79	0.87	39
Grey	0.89	0.85	0.87	474
Orange	0.88	0.75	0.81	28
Red	0.96	0.98	0.97	305
Silver	0.89	0.93	0.91	389
Unlisted	0.52	0.21	0.30	76
White	0.95	0.98	0.96	470
Yellow	0.94	0.94	0.94	33
accuracy			0.91	3029
macro avg	0.87	0.82	0.84	3029
weighted avg	0.90	0.91	0.90	3029

Лучше всего себя показала EfficientNet-B0, что говорит о том, что при грамотном подборе гиперпараметров, на небольшом датасете относительно легкая модель с меньшим числом параметров может оказаться точнее тяжелой. В нашем случае разница составила чуть меньше чем 2%.