

Klasyfikacja obrazów

Michał Kłos 7.02.2026

Zbiór danych

5 klas:

1. Road
2. Hybrid
3. MTB
4. Fold
5. Cargo



Zbiór danych

- 2 zbiory:
 - Real - 5 klas x 20 obrazków
 - Synthetic - 5 klas x 20 obrazków



Augmentacje

A1 - Baseline

- Zmiana rozmiaru (224, 224) + normalizacja

A2 - Geometria

- Przekształcenia przestrzenne (kadrowanie, odbicie, obrót)

A3 - Kolor

- Nasycenie, jitter, skala szarości

A4 - Połączzone

- Kombinacja wcześniejszych augmentacji

Warianty danych

E₁ - Pełny zbiór (Real + Synthetic)

E₂ - Tylko syntetyczne

E₃ - Tylko prawdziwe

E₄ - Zbilansowane dane

! Zbiorem testowym jest zawsze ta sama kolekcja
20 prawdziwych obrazków - po 4 na klasę !

Architektury modeli

1. MobileNetV2 - lekki i dopasowany do mniejszych zbiorów
 - ~3,4M parametrów
 - Szybkość i możliwość skutecznego uczenia na CPU
2. ResNet18 - połączenia rezydualne
 - ~18M parametrów
 - Lepsze rozwiązanie niż VGG dla takiego rozmiaru datasetu

Dodatkowo modele posiadają wagi przetrenowane na ImageNet

Wykonanie eksperymentów

Konfiguracja funkcji została zoptymalizowana pod mały zbiór danych

```
EXPERIMENT_CONFIG = {  
    'num_classes': 5,  
    'num_epochs': 30,  
    'batch_size': 16, # Small batch for small dataset  
    'learning_rate_scratch': 1e-3, # Higher LR for training from scratch  
    'learning_rate_finetune': 1e-4, # Lower LR for fine-tuning pretrained  
    'weight_decay': 1e-4,  
    'patience': 7, # Early stopping patience  
}
```

2 Architektury x 4 warianty danych x 4 augmentacje = 32 eksperymenty

Plus dodatkowe porównanie z modelem przetrenowanym na ImageNet

Przykładowe wyniki

```
=====
Experiment: ResNet18_scratch | E1_full | A3_color
Device: cuda | LR: 0.001 | Pretrained: False
=====
```

```
Data: train=120, val=20, test=20
```

```
Test Results (on FIXED test set of 20 images):
```

```
Accuracy: 55.00%
Precision: 55.33%
Recall: 55.00%
F1 Score: 53.65%
Time: 165.5s
```

```
=====
Experiment: MobileNetV2_scratch | E1_full | A2_geometric
Device: cuda | LR: 0.001 | Pretrained: False
=====
```

```
Data: train=120, val=20, test=20
```

```
Test Results (on FIXED test set of 20 images):
```

```
Accuracy: 30.00%
Precision: 33.33%
Recall: 30.00%
F1 Score: 25.00%
Time: 102.6s
```

Tabela wyników

	Arch	Data	Aug	Best Val Acc (%)	Test Acc (%)	Test Precision (%)	Test Recall (%)	Test F1 (%)
0	MobileNetV2	E1	A1	20.0	20.0	4.00	20.0	6.67
1	MobileNetV2	E1	A2	35.0	30.0	33.33	30.0	25.00
2	MobileNetV2	E1	A3	40.0	35.0	21.67	35.0	25.71
3	MobileNetV2	E1	A4	20.0	20.0	4.00	20.0	6.67
4	MobileNetV2	E2	A1	20.0	20.0	4.00	20.0	6.67
5	MobileNetV2	E2	A2	20.0	20.0	4.00	20.0	6.67
6	MobileNetV2	E2	A3	20.0	20.0	4.00	20.0	6.67
7	MobileNetV2	E2	A4	20.0	20.0	4.00	20.0	6.67
8	MobileNetV2	E3	A1	20.0	20.0	4.00	20.0	6.67
9	MobileNetV2	E3	A2	20.0	20.0	4.00	20.0	6.67
10	MobileNetV2	E3	A3	20.0	20.0	4.00	20.0	6.67
11	MobileNetV2	E3	A4	20.0	20.0	4.00	20.0	6.67
12	MobileNetV2	E4	A1	20.0	20.0	4.00	20.0	6.67
13	MobileNetV2	E4	A2	20.0	20.0	4.00	20.0	6.67
14	MobileNetV2	E4	A3	20.0	20.0	4.00	20.0	6.67
15	MobileNetV2	E4	A4	20.0	20.0	4.00	20.0	6.67
16	ResNet18	E1	A1	55.0	45.0	36.43	45.0	38.21
17	ResNet18	E1	A2	55.0	40.0	35.43	40.0	35.66
18	ResNet18	E1	A3	65.0	55.0	55.33	55.0	53.65
19	ResNet18	E1	A4	30.0	25.0	29.29	25.0	19.67
20	ResNet18	E2	A1	20.0	20.0	4.00	20.0	6.67
21	ResNet18	E2	A2	25.0	25.0	10.00	25.0	14.17
22	ResNet18	E2	A3	20.0	20.0	4.00	20.0	6.67
23	ResNet18	E2	A4	25.0	20.0	16.67	20.0	15.67
24	ResNet18	E3	A1	50.0	40.0	36.00	40.0	33.90
25	ResNet18	E3	A2	25.0	20.0	4.21	20.0	6.96
26	ResNet18	E3	A3	50.0	40.0	33.50	40.0	31.90
27	ResNet18	E3	A4	45.0	30.0	35.33	30.0	30.67
28	ResNet18	E4	A1	50.0	40.0	36.00	40.0	33.90
29	ResNet18	E4	A2	25.0	20.0	4.21	20.0	6.96
30	ResNet18	E4	A3	50.0	40.0	33.50	40.0	31.90
31	ResNet18	E4	A4	45.0	30.0	35.33	30.0	30.67

Wnioski

MobileNetV2 całkowicie zawodzi

- Test Accuracy = 20% w prawie wszystkich przypadkach
- To jest losowe zgadywanie dla 5 klas ($1/5 = 20\%$)
- Precyza = 4% i $F_1 = 6.67\%$ wskazują, że model niczego nie nauczył się

ResNet18 radził sobie nieco lepiej

- $E_1 + A_3$, $E_1 + A_1$,
- Prostsze warstwy konwolucyjne uczą się podstawowych cech (krawędzie, tekstury)
- Mniejsze ryzyko overfittingu niż MobileNetV2

Wpływ scenariuszy danych

Scenariusz	MobileNetV2	ResNet18	Interpretacja
E1	20-35%	35-55%	Więcej danych pomaga obu modelom
E2	20%	20-25%	Dane syntetyczne dają najsłabsze wyniki
E3	20%	20-40%	Mniejszy zbiór, ale Real > Synthetic

Wpływ augmentacji

Scenariusz	MobileNetV2	ResNet18	Interpretacja
A1	20%	40-45%	Bez augmentacji ryzyko overfit
A2	20-30%	20-40%	Brak jednoznacznej konkluzji
A3	20-35%	40-55%	Najlepsza dla ResNet18
A4	20%	25-30%	Zbyt dużo augmentacji

Modele pre-trained

```
=====
Experiment: ResNet18_pretrained | E1_full | A3_color
Device: cuda | LR: 0.0001 | Pretrained: True
=====
Data: train=120, val=20, test=20
```

Test Results (on FIXED test set of 20 images):
Accuracy: 95.00%
Precision: 96.00%
Recall: 95.00%
F1 Score: 94.92%
Time: 95.3s

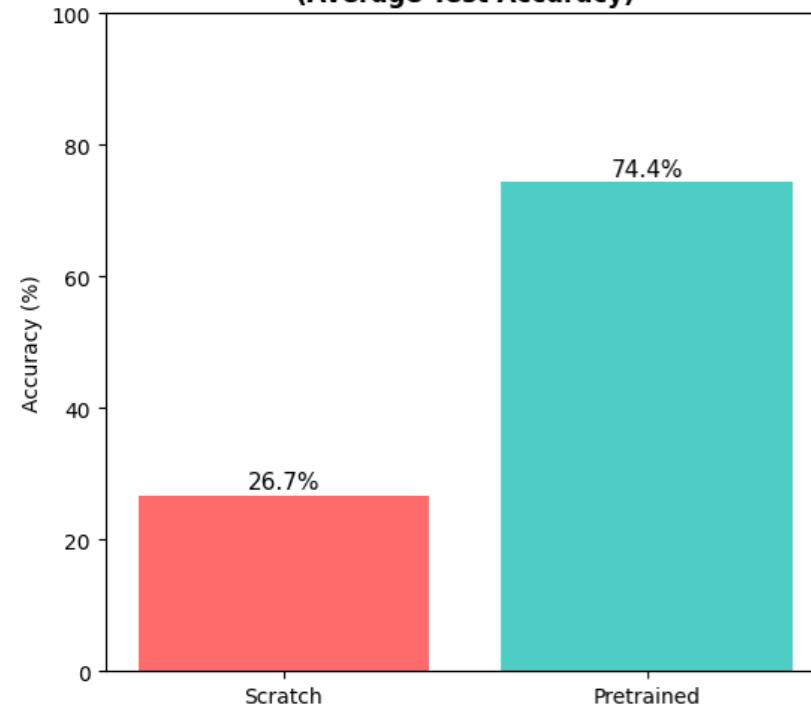
```
=====
Experiment: MobileNetV2_pretrained | E1_full | A3_color
Device: cuda | LR: 0.0001 | Pretrained: True
=====
Data: train=120, val=20, test=20
```

Test Results (on FIXED test set of 20 images):
Accuracy: 70.00%
Precision: 72.00%
Recall: 70.00%
F1 Score: 70.48%
Time: 75.5s

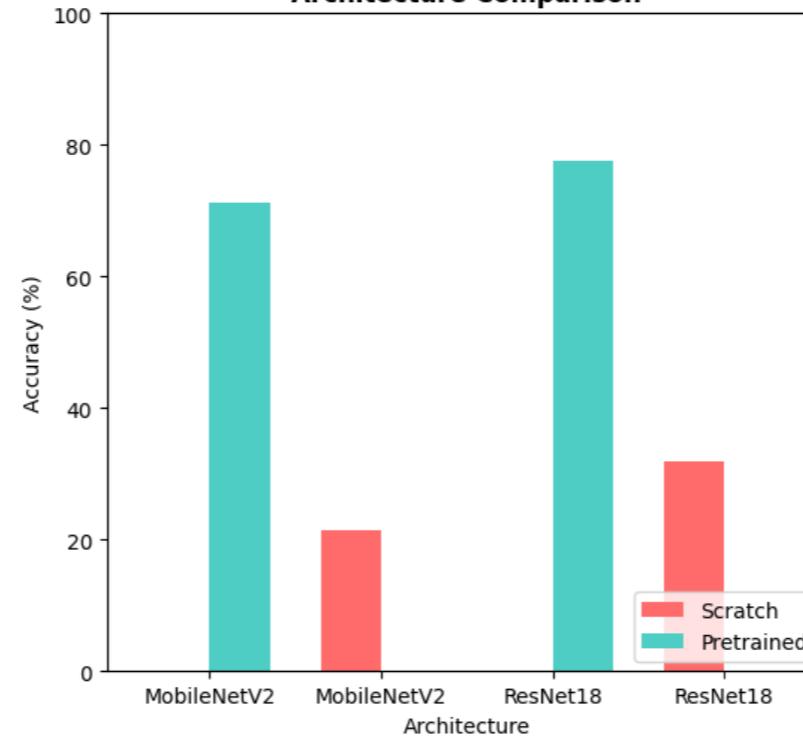
Tabela wyników (fine tuned)

	Arch	Data	Aug	Best Val Acc (%)	Test Acc (%)	Test Precision (%)	Test Recall (%)	Test F1 (%)
0	MobileNetV2	E1	A1	85.0	75.0	77.00	75.0	75.48
1	MobileNetV2	E1	A2	95.0	80.0	82.00	80.0	80.48
2	MobileNetV2	E1	A3	85.0	70.0	72.00	70.0	70.48
3	MobileNetV2	E1	A4	90.0	85.0	85.33	85.0	84.13
4	MobileNetV2	E2	A1	40.0	50.0	54.33	50.0	51.35
5	MobileNetV2	E2	A2	40.0	55.0	52.50	55.0	52.14
6	MobileNetV2	E2	A3	50.0	50.0	55.90	50.0	50.59
7	MobileNetV2	E2	A4	50.0	55.0	50.57	55.0	51.39
8	MobileNetV2	E3	A1	80.0	80.0	82.00	80.0	80.48
9	MobileNetV2	E3	A2	80.0	65.0	81.33	65.0	66.67
10	MobileNetV2	E3	A3	80.0	80.0	82.00	80.0	80.48
11	MobileNetV2	E3	A4	95.0	85.0	89.33	85.0	84.25
12	MobileNetV2	E4	A1	80.0	80.0	82.00	80.0	80.48
13	MobileNetV2	E4	A2	80.0	65.0	81.33	65.0	66.67
14	MobileNetV2	E4	A3	80.0	80.0	82.00	80.0	80.48
15	MobileNetV2	E4	A4	95.0	85.0	89.33	85.0	84.25
16	ResNet18	E1	A1	90.0	90.0	93.33	90.0	90.29
17	ResNet18	E1	A2	80.0	80.0	86.00	80.0	80.25
18	ResNet18	E1	A3	95.0	95.0	96.00	95.0	94.92
19	ResNet18	E1	A4	80.0	80.0	86.00	80.0	80.25
20	ResNet18	E2	A1	60.0	65.0	64.33	65.0	64.21
21	ResNet18	E2	A2	55.0	60.0	60.00	60.0	58.67
22	ResNet18	E2	A	70.0	75.0	82.00	75.0	75.81
23	ResNet18	E2	A4	55.0	55.0	51.71	55.0	52.19
24	ResNet18	E3	A1	85.0	90.0	93.33	90.0	90.29
25	ResNet18	E3	A2	85.0	95.0	96.00	95.0	94.92
26	ResNet18	E3	A3	85.0	80.0	82.00	80.0	80.48
27	ResNet18	E3	A4	70.0	55.0	60.00	55.0	49.90
28	ResNet18	E4	A1	85.0	90.0	93.33	90.0	90.29
29	ResNet18	E4	A	85.0	95.0	96.00	95.0	94.92
30	ResNet18	E4	A3	85.0	80.0	82.00	80.0	80.48
31	ResNet18	E4	A4	70.0	55.0	60.00	55.0	49.90

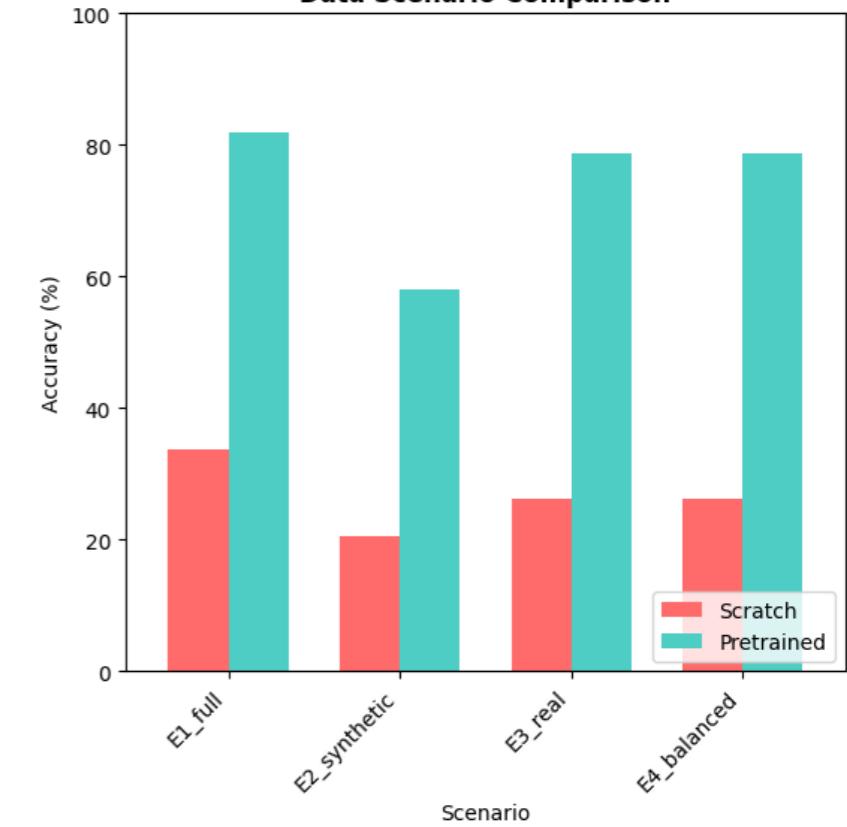
**Scratch vs Pretrained
(Average Test Accuracy)**



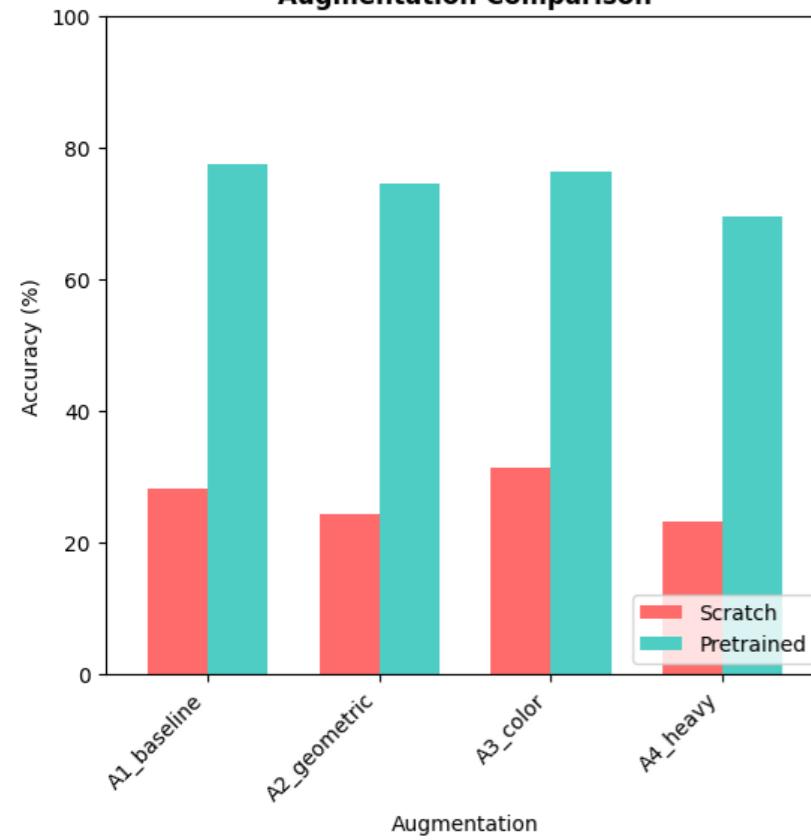
Architecture Comparison



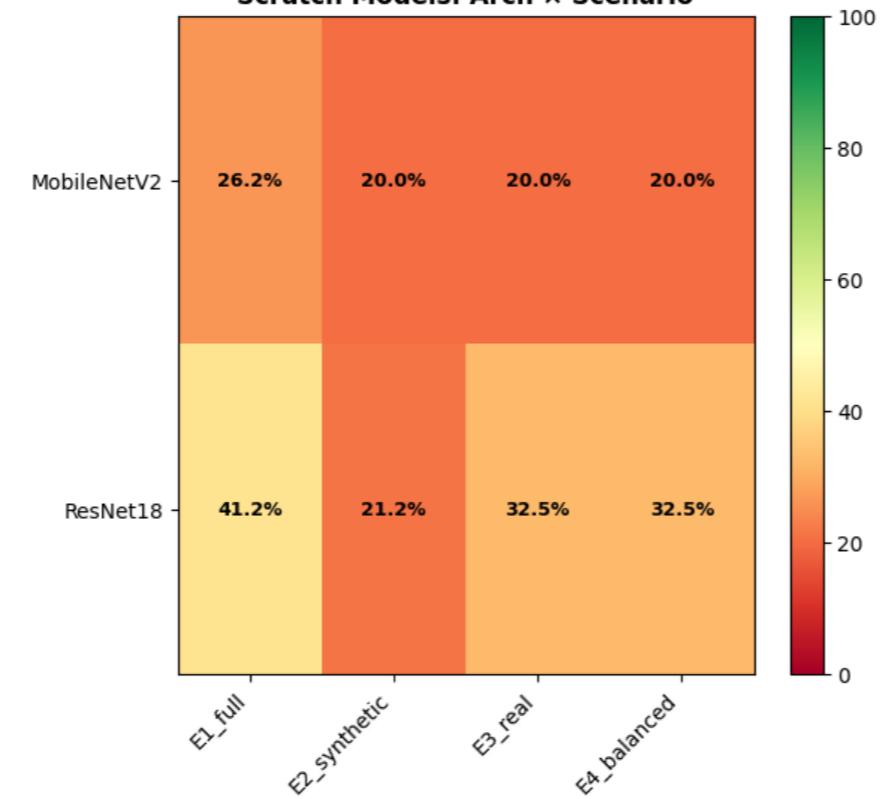
Data Scenario Comparison



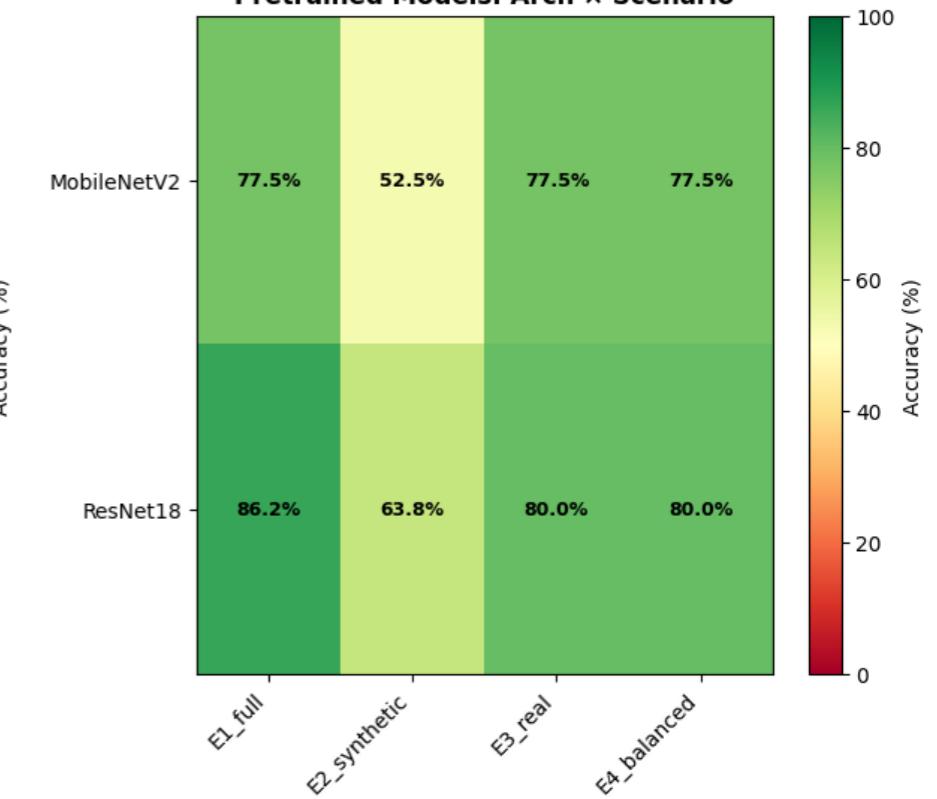
Augmentation Comparison



Scratch Models: Arch x Scenario



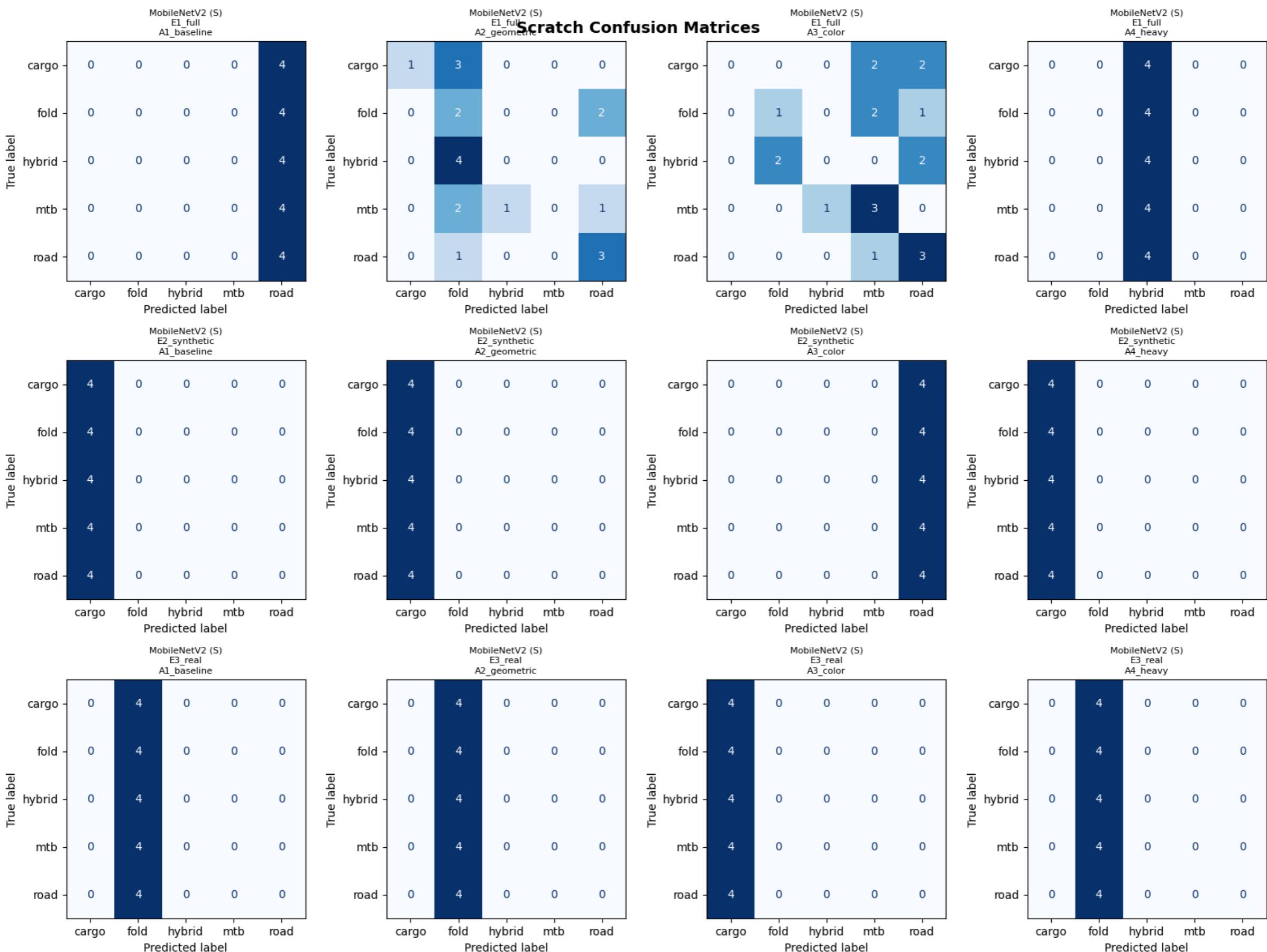
Pretrained Models: Arch x Scenario

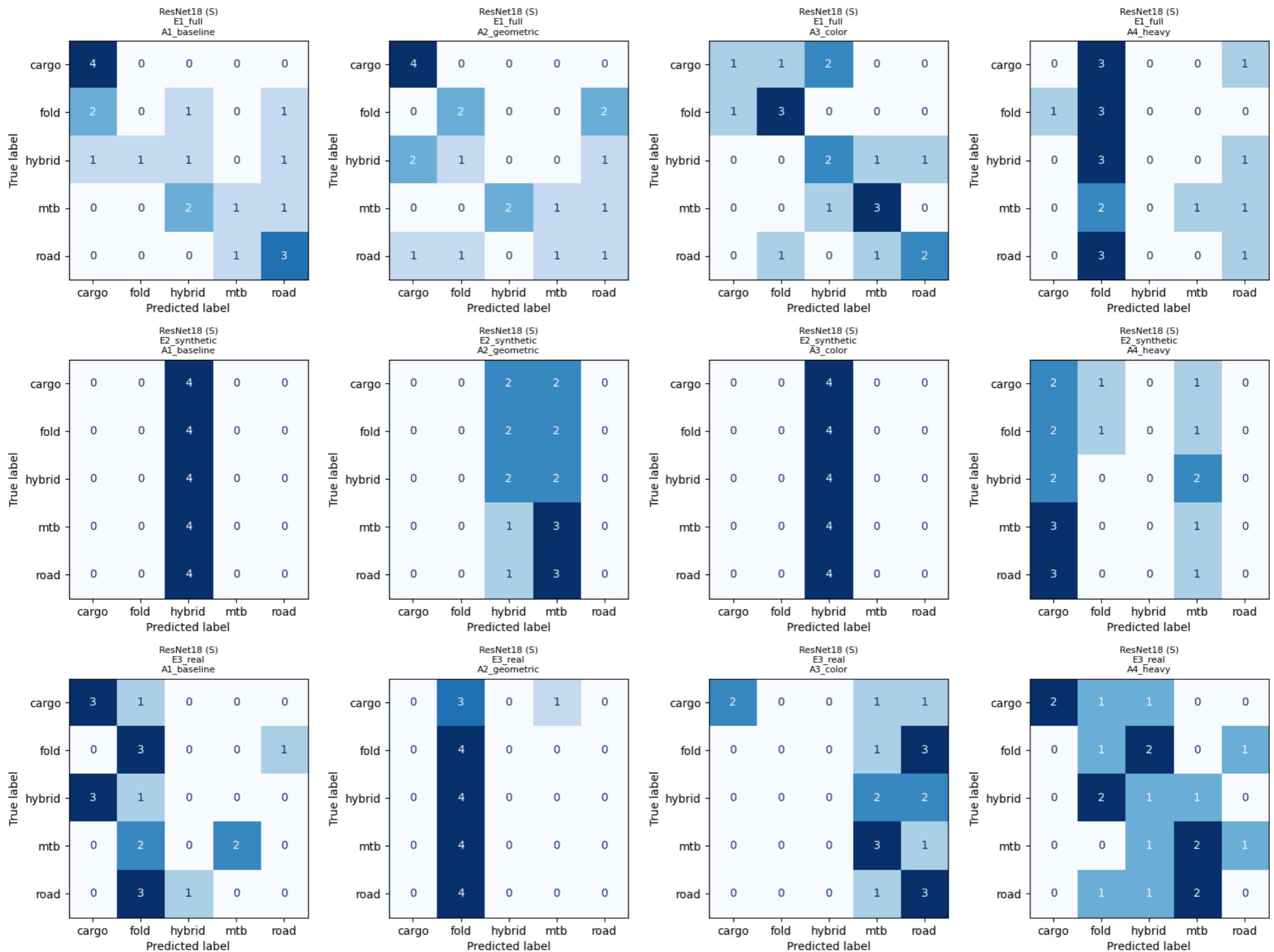


Tablice pomyłek

Modele trenowane od zera

Scratch Confusion Matrices

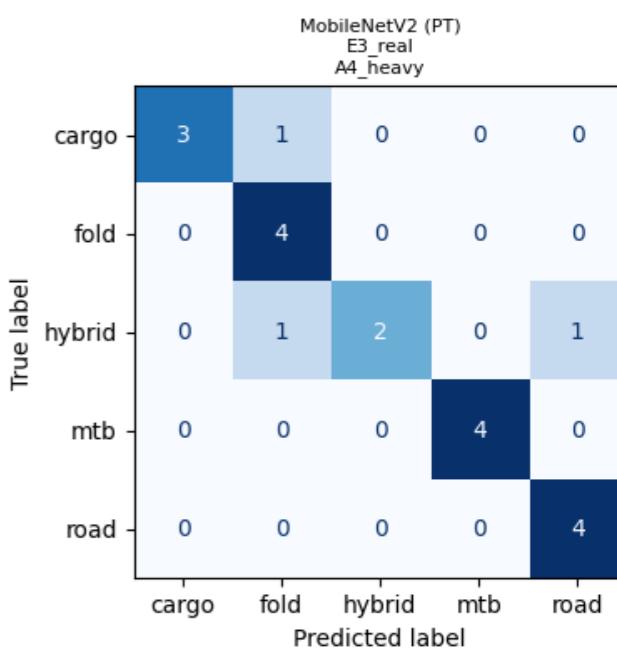
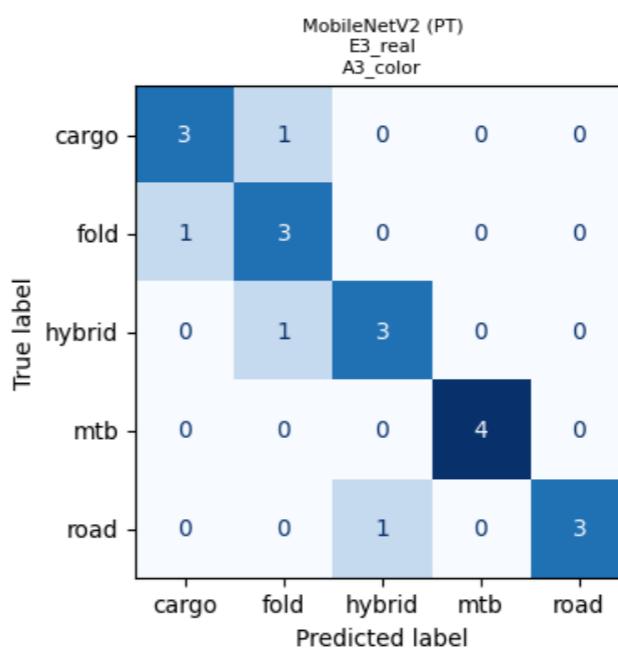
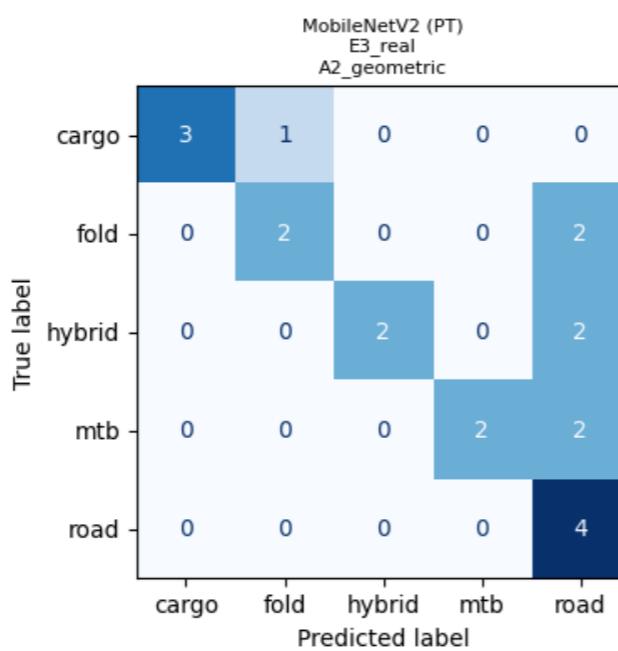
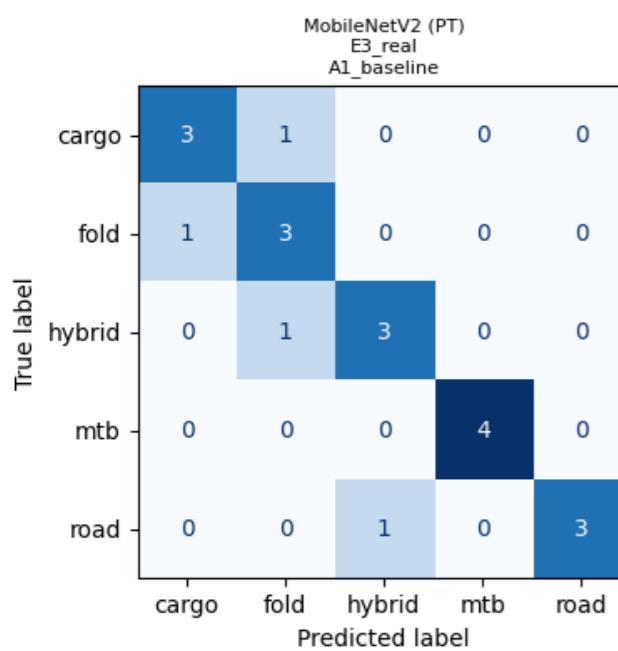
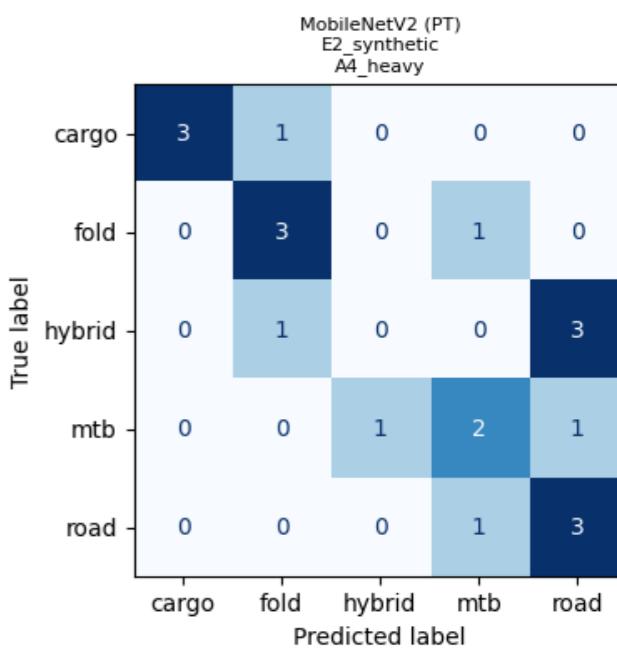
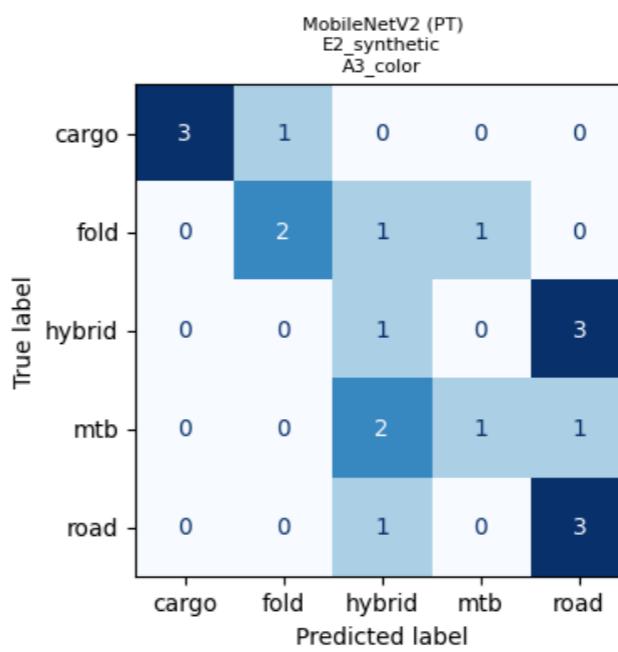
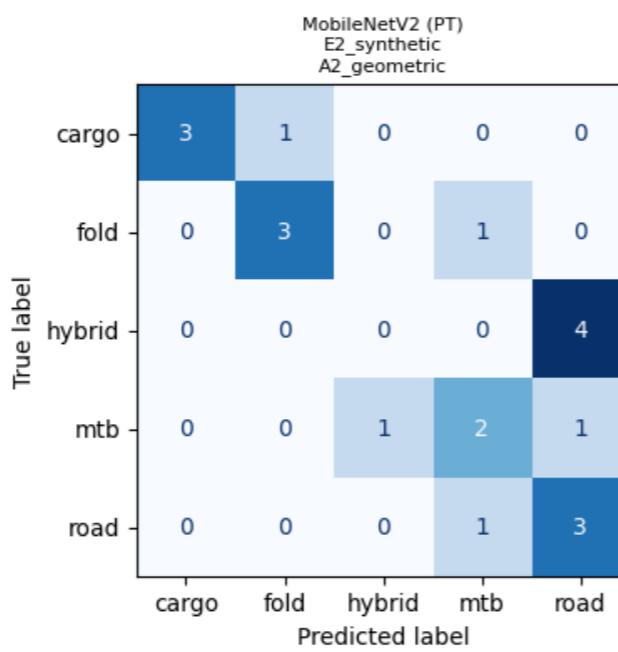
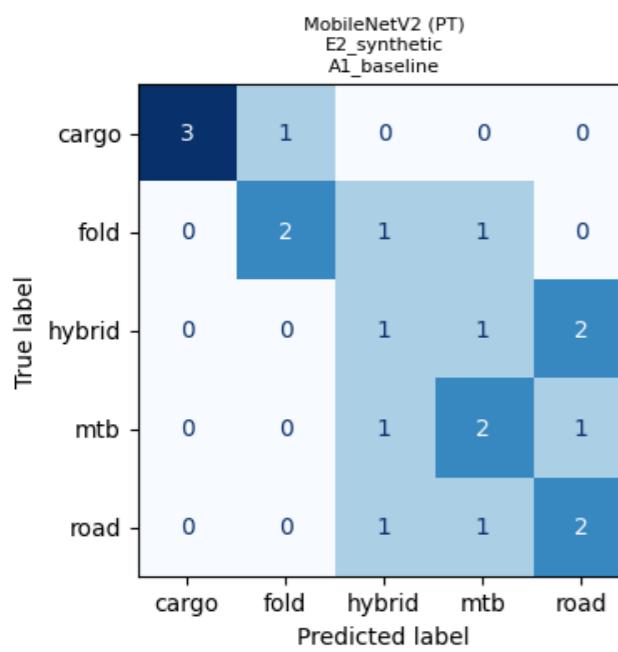
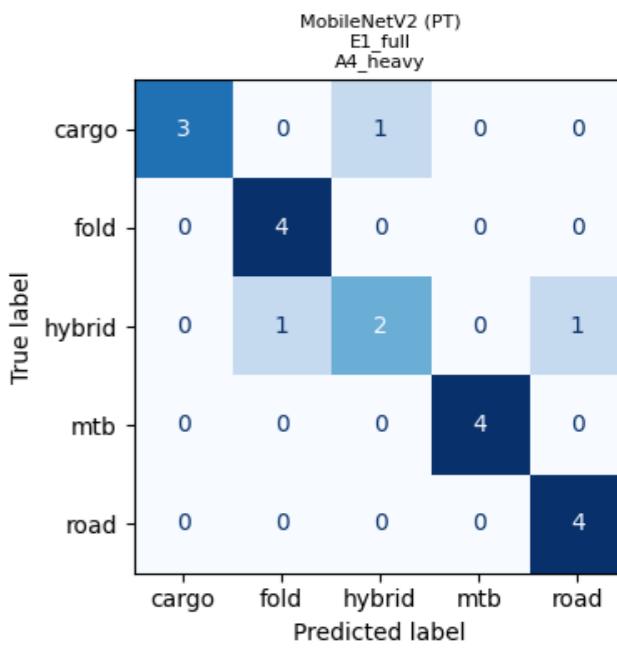
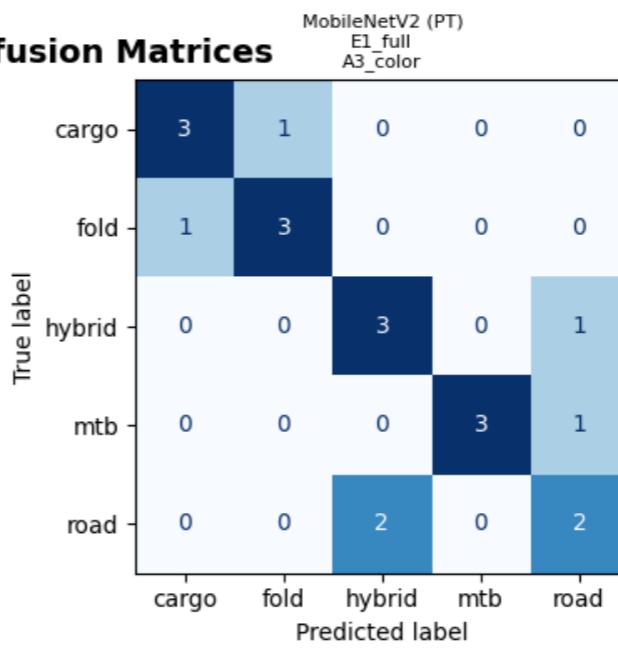
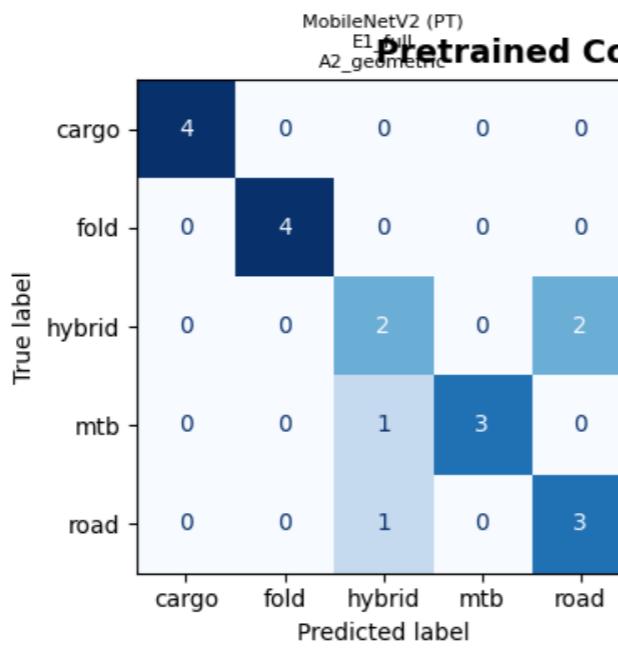
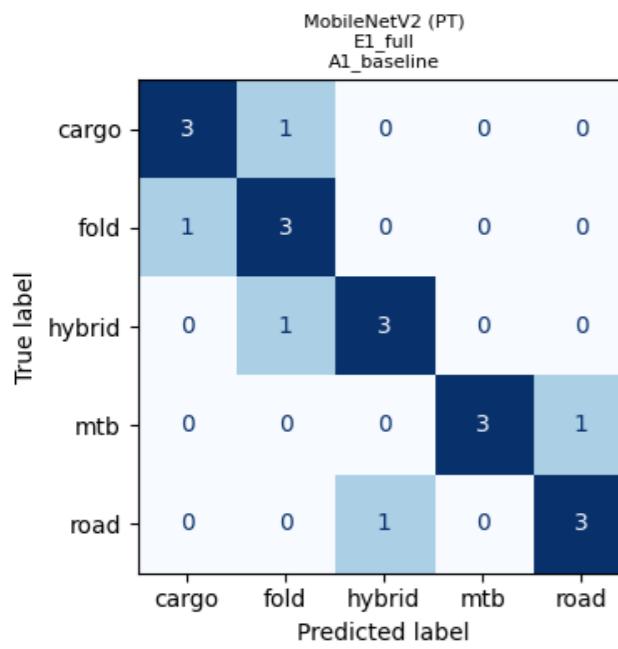


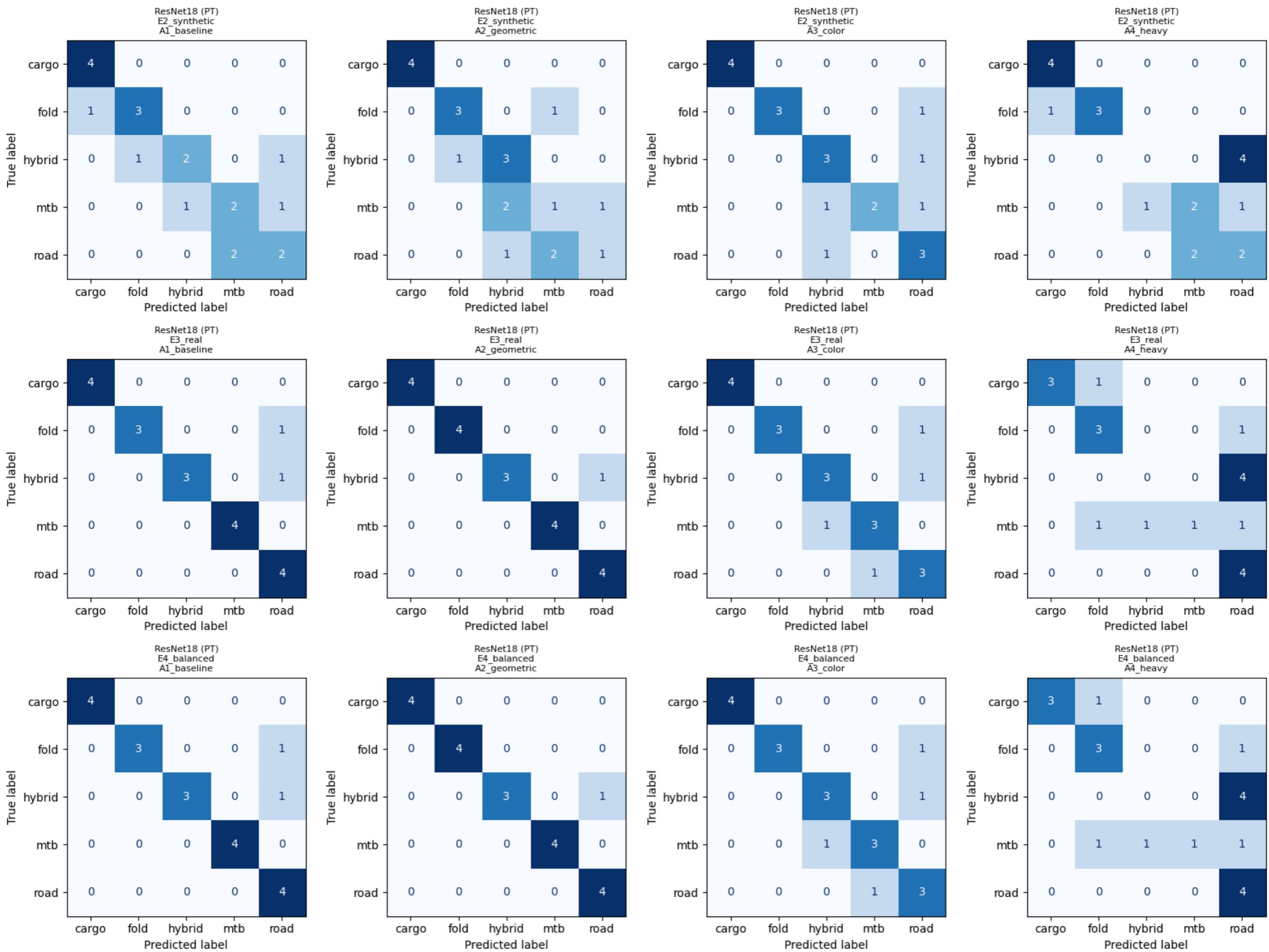


Tablice pomyłek

Modele dostrajane na bazie ImageNet

Pretrained Confusion Matrices





Koniec