

# TRASH TEST



# INTRODUZIONE E NOMENCLATURA

1. DATASET DI TRAINING
2. DATASET DI TESTING
3. ESPERIMENTI

# Dataset cleaned: C

È il dataset di partenza, usato spesso in letteratura per il task di waste recognition. Presenta immagini a sfondo neutro, per massimizzare la distinzione dei contorni degli oggetti.

**Cardboard**



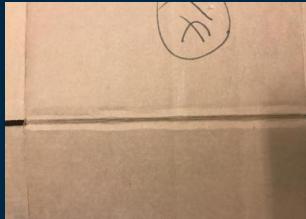
**Compost**



**Glass**



**Metal**



# Dataset cleaned: C

È il dataset di partenza, usato spesso in letteratura per il task di waste recognition. Presenta immagini a sfondo neutro, per massimizzare la distinzione dei contorni degli oggetti.

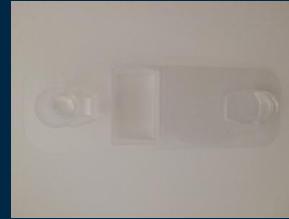
Paper



Plastic



Trash



# Dataset con background aumentato: BG

Le immagini del dataset C sono state private del background neutro ed è stato aggiunto un background "più realistico". Alcuni background sono stati presi da internet, altri sono stati fotografati da me per riprendere sfondi che troveremo nel testset realistico più avanti.





# Dataset con background aumentato: BG

Le immagini del dataset C sono state private del background neutro ed è stato aggiunto un background “più realistico”. Alcuni background sono stati presi da internet, altri sono stati fotografati da me per riprendere sfondi che troveremo nel testset realistico più avanti.

**Cardboard**



Buon  
overlay

**Compost**



**Glass**



**Metal**



Pessimo  
overlay



# Dataset con background aumentato: BG

Le immagini del dataset C sono state private del background neutro ed è stato aggiunto un background “più realistico”. Alcuni background sono stati presi da internet, altri sono stati fotografati da me per riprendere sfondi che troveremo nel testset realistico più avanti.

**Paper**



**Buon  
overlay**

**Plastic**



**Trash**



**Pessimo  
overlay**



# Dataset C+BG

Il dataset composto dalle immagini dei primi due, dunque per ogni immagine a sfondo neutro è presente una sua copia a sfondo aumentato.

Distribuzione dataset C (e BG)

Category	Entries	Percentage
Cardboard	403	15
Compost	177	6
Glass	501	18
Metal	410	15
Paper	594	22
Plastic	482	17
Trash	184	7
<b>Total</b>	2751	100

Distribuzione dataset C + BG

Category	Entries	Percentage
Cardboard	806	15
Compost	354	6
Glass	1002	18
Metal	820	15
Paper	1188	22
Plastic	964	17
Trash	368	7
<b>Total</b>	5502	100

# INTRODUZIONE E NOMENCLATURA

1. DATASET DI TRAINING
2. DATASET DI TESTING
3. ESPERIMENTI

# Test split

Quando si fanno esperimenti su un certo modello, per test split si intende immagini dello stesso tipo del dataset di training, ma ovviamente mai viste in questa fase.

Gli split dei dataset hanno tutti ratio 70%, 13% e 17% rispettivamente per train, validation e test.

# Realistic test set

Insieme di foto scattate a rifiuti in diversi contesti (indoor e outdoor) creato ad hoc per testare i modelli su immagini più “realistiche” e con sfondo non neutro.

**Cardboard**



**Compost**



**Glass**



**Metal**



# Realistic test set

Insieme di foto scattate a rifiuti in diversi contesti (indoor e outdoor) creato ad hoc per testare i modelli su immagini più “realistiche” e con sfondo non neutro.

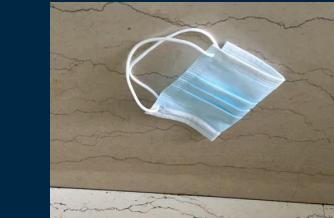
**Paper**



**Plastic**



**Trash**



**Distribuzione realistic test set**

Category	Entries	Percentage
Cardboard	15	13
Compost	19	17
Glass	11	10
Metal	11	10
Paper	15	13
Plastic	26	23
Trash	16	14
Total	114	100

# INTRODUZIONE E NOMENCLATURA

1. DATASET DI TRAINING
2. DATASET DI TESTING
3. ESPERIMENTI

# Nomenclatura dei modelli

- Modello A: sequenziale, più semplice, basato su AlexNet
- Modello B: non sequenziale, con skip connections, più complesso, basato su InceptionResNet V2

# Nomenclatura dei test set

- T1: immagini con background neutro ( C )
- T2: immagini realistiche ( realistic )
- T3: immagini con background aumentato ( BG )
- T4: immagini con background aumentati e non (C + BG)

# Nomenclatura degli esperimenti

{modello}\_{training}\_{testset}

Esempio:

report\_A\_BG\_T4

Report del **modello A**, con training fatto su dataset composto da sole immagini con **background aumentato** e test su immagini con **background aumentato e non.**

# ESPERIMENTI



# report\_A\_C\_T1

	precision	recall	f1-score	support
0	0.79	0.72	0.76	69
1	0.70	0.84	0.76	31
2	0.62	0.63	0.62	86
3	0.66	0.76	0.71	70
4	0.81	0.85	0.83	102
5	0.67	0.58	0.62	83
6	0.56	0.45	0.50	33
accuracy			0.70	474
macro avg	0.69	0.69	0.69	474
weighted avg	0.70	0.70	0.70	474

# report\_A\_C\_T2

	precision	recall	f1-score	support
0	0.10	0.20	0.13	15
1	0.00	0.00	0.00	19
2	0.17	0.18	0.17	11
3	0.11	0.27	0.16	11
4	0.14	0.27	0.18	15
5	0.20	0.12	0.15	26
6	0.00	0.00	0.00	16
accuracy				0.13
macro avg			0.10	113
weighted avg			0.10	113

# report\_A\_C\_T3

## Hyperparameters:

-epoch: 20

-batch-size: 32

-preprocessing: vgg16

-lr: 1e-4

	precision	recall	f1-score	support
0	0.24	0.33	0.28	69
1	0.00	0.00	0.00	31
2	0.34	0.22	0.27	86
3	0.20	0.61	0.30	70
4	0.30	0.30	0.30	102
5	0.40	0.12	0.19	83
6	0.00	0.00	0.00	50
accuracy			0.26	491
macro avg	0.21	0.23	0.19	491
weighted avg	0.25	0.26	0.22	491

- 0. Cardboard
- 1. Compost
- 2. Glass
- 3. Metal
- 4. Paper
- 5. Plastic
- 6. Trash

T1: C  
T2: Real  
T3: BG  
T4: C+BG

# report\_A\_BG\_T1

	precision	recall	f1-score	support
0	0.40	0.55	0.46	69
1	0.59	0.42	0.49	31
2	0.40	0.37	0.38	86
3	0.46	0.51	0.48	70
4	0.57	0.26	0.36	102
5	0.39	0.55	0.46	83
6	0.27	0.24	0.25	33
accuracy			0.42	474
macro avg	0.44	0.42	0.41	474
weighted avg	0.45	0.42	0.42	474

# report\_A\_BG\_T2

	precision	recall	f1-score	support
0	0.10	0.20	0.14	15
1	0.00	0.00	0.00	19
2	0.11	0.09	0.10	11
3	0.10	0.18	0.12	11
4	0.07	0.13	0.09	15
5	0.26	0.23	0.24	26
6	0.00	0.00	0.00	16
accuracy				0.12
macro avg			0.09	0.12
weighted avg			0.10	0.11

# report\_A\_BG\_T3

## Hyperparameters:

-epoch: 20

-batch-size: 32

-preprocessing: vgg16

-lr: 1e-4

	precision	recall	f1-score	support
0	0.36	0.46	0.41	69
1	0.60	0.68	0.64	31
2	0.30	0.19	0.23	86
3	0.37	0.40	0.39	70
4	0.50	0.32	0.39	102
5	0.40	0.52	0.45	83
6	0.29	0.42	0.35	33
accuracy			0.39	474
macro avg	0.40	0.43	0.41	474
weighted avg	0.40	0.39	0.39	474

- 0. Cardboard
- 1. Compost
- 2. Glass
- 3. Metal
- 4. Paper
- 5. Plastic
- 6. Trash

T1: C  
T2: Real  
T3: BG  
T4: C+BG

# report\_A\_C+BG\_T1

	precision	recall	f1-score	support
0	0.65	0.81	0.72	69
1	0.71	0.77	0.74	31
2	0.53	0.52	0.53	86
3	0.71	0.67	0.69	70
4	0.74	0.71	0.72	102
5	0.61	0.59	0.60	83
6	0.50	0.39	0.44	33
accuracy			0.65	474
macro avg	0.64	0.64	0.63	474
weighted avg	0.64	0.65	0.64	474

# report\_A\_C+BG\_T2\*\*

	precision	recall	f1-score	support
0	0.09	0.20	0.13	15
1	0.00	0.00	0.00	19
2	0.21	0.27	0.24	11
3	0.25	0.36	0.30	11
4	0.07	0.13	0.10	15
5	0.39	0.27	0.32	26
6	0.00	0.00	0.00	16
accuracy				0.17
macro avg			0.15	113
weighted avg			0.16	113

# report A\_C+BG\_T4

## Hyperparameters:

-epoch: 20

-batch-size: 32

-preprocessing: vgg16

-lr: 1e-4

	precision	recall	f1-score	support
0	0.53	0.70	0.60	138
1	0.65	0.66	0.66	62
2	0.46	0.45	0.46	172
3	0.60	0.56	0.58	140
4	0.65	0.61	0.63	204
5	0.53	0.49	0.51	166
6	0.37	0.35	0.36	66
accuracy			0.55	948
macro avg	0.54	0.55	0.54	948
weighted avg	0.55	0.55	0.55	948

- 0. Cardboard
- 1. Compost
- 2. Glass
- 3. Metal
- 4. Paper
- 5. Plastic
- 6. Trash

T1: C  
T2: Real  
T3: BG  
T4: C+BG

# report\_B\_C\_T1

	precision	recall	f1-score	support
0	0.97	0.87	0.92	69
1	0.97	0.90	0.93	31
2	0.86	0.90	0.88	86
3	0.89	0.90	0.89	70
4	0.84	0.93	0.88	102
5	0.88	0.78	0.83	83
6	0.77	0.82	0.79	33
accuracy			0.88	474
macro avg	0.88	0.87	0.87	474
weighted avg	0.88	0.88	0.88	474

# report\_B\_C\_T2

	precision	recall	f1-score	support
0	0.47	0.60	0.53	15
1	1.00	0.47	0.64	19
2	0.62	0.73	0.67	11
3	0.50	0.36	0.42	11
4	0.26	0.67	0.38	15
5	0.50	0.19	0.28	26
6	0.12	0.12	0.12	16
accuracy				0.42
macro avg			0.50	0.45
weighted avg			0.51	0.42

# report\_B\_C\_T3

## Hyperparameters:

-epoch: 80

-batch-size: 32

-preprocessing: IRSV2

-lr: 1e-4

	precision	recall	f1-score	support
0	0.34	0.54	0.41	69
1	0.17	0.06	0.09	31
2	0.63	0.44	0.52	86
3	0.58	0.61	0.60	70
4	0.43	0.62	0.51	102
5	0.72	0.46	0.56	83
6	0.49	0.34	0.40	50
accuracy			0.48	491
macro avg	0.48	0.44	0.44	491
weighted avg	0.51	0.48	0.48	491

- 0. Cardboard
- 1. Compost
- 2. Glass
- 3. Metal
- 4. Paper
- 5. Plastic
- 6. Trash

T1: C  
T2: Real  
T3: BG  
T4: C+BG

# report\_B\_BG\_T1

	precision	recall	f1-score	support
0	0.90	0.67	0.77	69
1	0.93	0.84	0.88	31
2	0.65	0.79	0.72	86
3	0.78	0.80	0.79	70
4	0.82	0.76	0.79	102
5	0.65	0.66	0.65	83
6	0.49	0.58	0.53	33
accuracy			0.73	474
macro avg	0.75	0.73	0.73	474
weighted avg	0.75	0.73	0.74	474

# report\_B\_BG\_T2

	precision	recall	f1-score	support
0	0.62	0.33	0.43	15
1	1.00	0.26	0.42	19
2	0.47	0.73	0.57	11
3	0.26	0.64	0.37	11
4	0.38	0.60	0.46	15
5	0.58	0.54	0.56	26
6	0.25	0.12	0.17	16
accuracy				0.44
macro avg			0.51	0.46
weighted avg			0.54	0.43

# report\_B\_BG\_T3

## Hyperparameters:

-epoch: 80

-batch-size: 32

-preprocessing: IRSV2

-lr: 1e-4

	precision	recall	f1-score	support
0	0.61	0.71	0.66	69
1	0.80	0.65	0.71	31
2	0.54	0.76	0.63	86
3	0.76	0.77	0.77	70
4	0.76	0.67	0.71	102
5	0.76	0.47	0.58	83
6	0.49	0.55	0.51	33
accuracy			0.66	474
macro avg	0.67	0.65	0.65	474
weighted avg	0.68	0.66	0.66	474

- 0. Cardboard
- 1. Compost
- 2. Glass
- 3. Metal
- 4. Paper
- 5. Plastic
- 6. Trash

T1: C  
T2: Real  
T3: BG  
T4: C+BG

# report\_B\_C+BG\_T1

	precision	recall	f1-score	support
0	0.93	0.93	0.93	69
1	0.97	0.94	0.95	31
2	0.85	0.91	0.88	86
3	0.88	0.93	0.90	70
4	0.89	0.91	0.90	102
5	0.83	0.76	0.79	83
6	0.69	0.61	0.65	33
accuracy			0.87	474
macro avg	0.86	0.85	0.86	474
weighted avg	0.87	0.87	0.87	474

# report\_B\_C+BG\_T2\*\*

	precision	recall	f1-score	support
0	0.53	0.67	0.59	15
1	1.00	0.47	0.64	19
2	0.67	0.73	0.70	11
3	0.53	0.91	0.67	11
4	0.41	0.60	0.49	15
5	0.62	0.62	0.62	26
6	0.33	0.12	0.18	16
accuracy				0.57
macro avg	0.58	0.59	0.55	113
weighted avg	0.60	0.57	0.55	113

# report B\_C+BG\_T4

## Hyperparameters:

-epoch: 80

-batch-size: 32

-preprocessing: IRSV2

-lr: 1e-4

	precision	recall	f1-score	support
0	0.77	0.82	0.79	138
1	0.89	0.79	0.84	62
2	0.75	0.77	0.76	172
3	0.77	0.86	0.81	140
4	0.78	0.83	0.81	204
5	0.79	0.70	0.74	166
6	0.72	0.52	0.60	66
accuracy			0.78	948
macro avg	0.78	0.76	0.77	948
weighted avg	0.78	0.78	0.77	948

- 0. Cardboard
- 1. Compost
- 2. Glass
- 3. Metal
- 4. Paper
- 5. Plastic
- 6. Trash

- T1: C
- T2: Real
- T3: BG
- T4: C+BG

# Osservazioni

Il risultato più interessante è che nel test realistico per entrambi i modelli aggiungere le immagini con background aumentato a quelle neutre in fase di training migliora le prestazioni, per il caso del modello B in modo considerevole.

report\_A\_C\_T2

	precision	recall	f1-score	support
0	0.10	0.20	0.13	15
1	0.00	0.00	0.00	19
2	0.17	0.18	0.17	11
3	0.11	0.27	0.16	11
4	0.14	0.27	0.18	15
5	0.20	0.12	0.15	26
6	0.00	0.00	0.00	16
accuracy			0.13	113
macro avg	0.10	0.15	0.11	113
weighted avg	0.10	0.13	0.11	113

report\_A\_C+BG\_T2\*\*

	precision	recall	f1-score	support
0	0.09	0.20	0.13	15
1	0.00	0.00	0.00	19
2	0.21	0.27	0.24	11
3	0.25	0.36	0.30	11
4	0.07	0.13	0.10	15
5	0.39	0.27	0.32	26
6	0.00	0.00	0.00	16
accuracy			0.17	113
macro avg	0.15	0.18	0.15	113
weighted avg	0.16	0.17	0.15	113

# Osservazioni

Il risultato più interessante è che nel test realistico per entrambi i modelli aggiungere le immagini con background aumentato a quelle neutre in fase di training migliora le prestazioni, per il caso del modello B in modo considerevole.

report\_B\_C\_T2

	precision	recall	f1-score	support
0	0.40	0.67	0.50	15
1	1.00	0.42	0.59	19
2	0.64	0.64	0.64	11
3	0.42	0.45	0.43	11
4	0.25	0.60	0.35	15
5	0.50	0.19	0.28	26
6	0.18	0.12	0.15	16
accuracy			0.41	113
macro avg	0.48	0.44	0.42	113
weighted avg	0.50	0.41	0.40	113

report\_B\_C+BG\_T2\*\*

	precision	recall	f1-score	support
0	0.53	0.67	0.59	15
1	1.00	0.47	0.64	19
2	0.67	0.73	0.70	11
3	0.53	0.91	0.67	11
4	0.41	0.60	0.49	15
5	0.62	0.62	0.62	26
6	0.33	0.12	0.18	16
accuracy			0.57	113
macro avg	0.58	0.59	0.55	113
weighted avg	0.60	0.57	0.55	113

# Osservazioni

È anche interessante notare che nel training con le sole immagini con background aumentato sembrano rimanere stabili nel modello B, e peggiorare leggermente nel modello A.

Un risultato simile vale per i test sulle immagini neutre con il training C+BG.

Il fatto di aver riscontrato risultati coerenti per entrambi i modelli sembra essere una conferma dell'efficacia di questa tecnica di background augmentation. Nello specifico, quando la tecnica non comporta miglioramenti evidenti, nel modello A abbiamo un leggero calo delle prestazioni, mentre nel modello B un leggero miglioramento. Quando invece il suo impatto è maggiore, nel modello A vediamo poco miglioramento e nel modello B un miglioramento sostanziale.

# Tabella di confronto

	C		BG		C+BG		
	A	B	A	B	A	B	
<b>TEST 1</b>							
Accuracy		0.70	0.88	0.42	0.73	0.65	0.87
Macro AVG	F1	0.69	0.87	0.41	0.73	0.63	0.86
Weighted AVG	F1	0.70	0.88	0.42	0.74	0.64	0.87
<b>TEST 2</b>							
Accuracy		0.13	0.42	0.12	0.44	0.17	0.57
Macro AVG	F1	0.11	0.43	0.10	0.43	0.15	0.55
Weighted AVG	F1	0.11	0.42	0.11	0.43	0.15	0.55
<b>TEST 3</b>							
Accuracy		0.26	0.48	0.39	0.66	\	\
Macro AVG	F1	0.19	0.44	0.41	0.65	\	\
Weighted AVG	F1	0.22	0.48	0.39	0.66	\	\
<b>TEST 4</b>							
Accuracy		\	\	\	\	0.55	0.78
Macro AVG	F1	\	\	\	\	0.54	0.77
Weighted AVG	F1	\	\	\	\	0.55	0.77

**TEST 1: C**  
**TEST 2: Real**  
**TEST 3: BG**  
**TEST 4: C+BG**