Modulo E

Case Study (Esperimenti di Apprendimento)

Modalità

Divisione in 8 gruppi da 3 persone + 1 gruppo da 2 persone

Possibili case study:

- DEEP LEARNING:
    - Image classification con reti preaddestrate e transfer learning (e fine tuning)
    - Scelta di un dataset e confronto di diverse reti
- MACHINE LEARNING
    - Confronto di algoritmi di ML classici su uno o più dataset tabellari

Presentazione finale di max 10 minuti per gruppo: 16 giugno ore 16

Modalità

- DEEP LEARNING:
    - Si suggerisce di testare **più** reti preaddestrate sul dataset scelto
- MACHINE LEARNING
    - nel caso sia possibile applicare sia classification sia clustering si può scegliere se implementare uno o entrambi i task

Datasets for Image Classification

1 - Cartelli stradali

Dataset scaricabile al link:
https://sid.erda.dk/public/archives/ff17dc924eba88d5d01a807357d6614c/FullIJCNN2013.zip

- 43 classi
- Più di 50,000 immagini in totale

2 - Classificazione arance

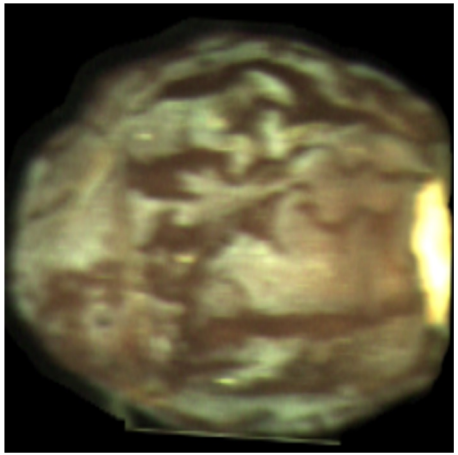
- 2 classi: buone, marce
- Dataset sbilanciato: 47 buone, 12 marce
- https://drive.google.com/file/d/1mFCLEOv75R3Se10eXB2n4OVuJrM-uliX

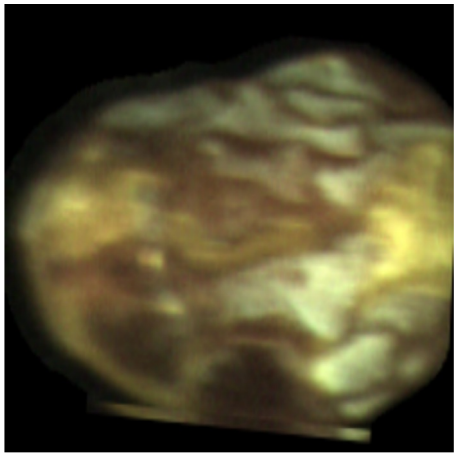
- Dentro l'archivio scaricabile c'è un file che riporta le classi di ogni immagine.

3 - Classificazione datteri

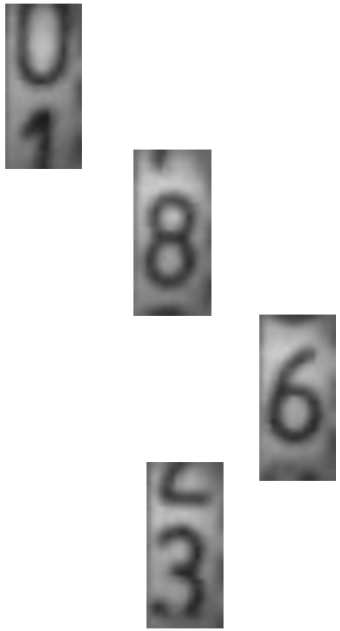
- 2 classi: buoni, non buoni
- Dataset sbilanciato: 100 buone, 140 marce
- https://drive.google.com/file/d/1fUoB3iImXLMhRrMH_MqFSvg9ce1sHG36

- La classe di ogni immagine è specificata nel nome del file: “date_good” per datteri buoni e “date_looseskin” per datteri non buoni





4 - Classificazione digit da contatori



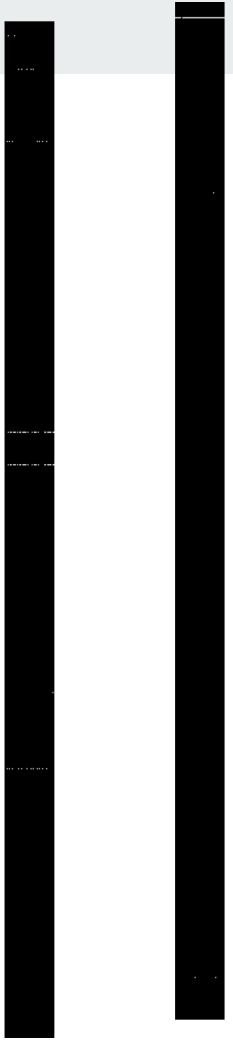
- 10 classi: numeri da 0 a 9
- Dataset bilanciato
- https://drive.google.com/file/d/1Sj-z7tqL3d3CFuvrPcuzFEhptlJWAQPv

- Le immagini sono divise in cartelle corrispondenti alla classe di appartenenza

5 - Classificazione dati genetici

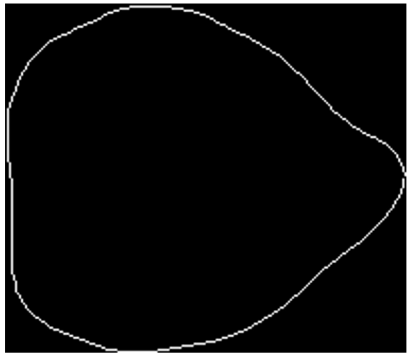
- 2 classi: mutazioni da processi casuali (neutrali) e da selezione naturale
- Dataset bilanciato
- https://drive.google.com/file/d/1KEUGRL4X4vTWYBHE7sayCIR8krJunORg

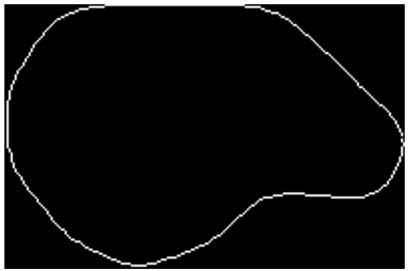
- Le immagini sono divise in cartelle corrispondenti alla classe di appartenenza



6 - Classificazione pere in base alla forma

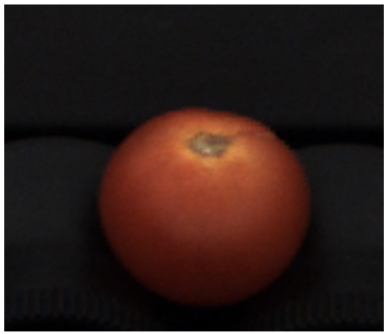
- 5 classi: asimmetriche, allungate, fat bottom, arrotondate, buone
- Dataset sbilanciato
- https://drive.google.com/file/d/1PnrYCDQ0tQ8IyfcQ0hoLimy4iBp7Y_EJ
- Le classi sono specificate nel file classes.csv

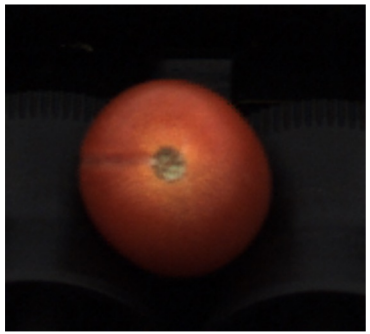




7 - Classificazione pomodorini

- 2 classi: buoni, crepati
- Dataset bilanciato (530-560 per classe)
- https://drive.google.com/file/d/1JDVTj0V26AMFenQ69KhWkmX41kPoTqAz
- Le immagini sono divise in cartelle corrispondenti alla classe di appartenenza





8 - Sasso carta forbice

Dataset scaricabile dal seguente link:

https://public.roboflow.com/classification/rock-paper-scissors

- 3 classi: sasso, carta, forbice
- Dataset bilanciato (840 immagini per classe)

9 - Tensorflow flowers

Dataset scaricabile a questo link:
https://storage.googleapis.com/download.tensorflow.org/example_images/flower_photos.tgz

- 5 classi: daisy, dandelion, rose, sunflower, tulip
- Dataset leggermente sbilanciato

10 – Classificazione Cat/Dog (Microsoft)

Scaricabile al seguente link: https://www.microsoft.com/en-us/download/confirmation.aspx?id=54765

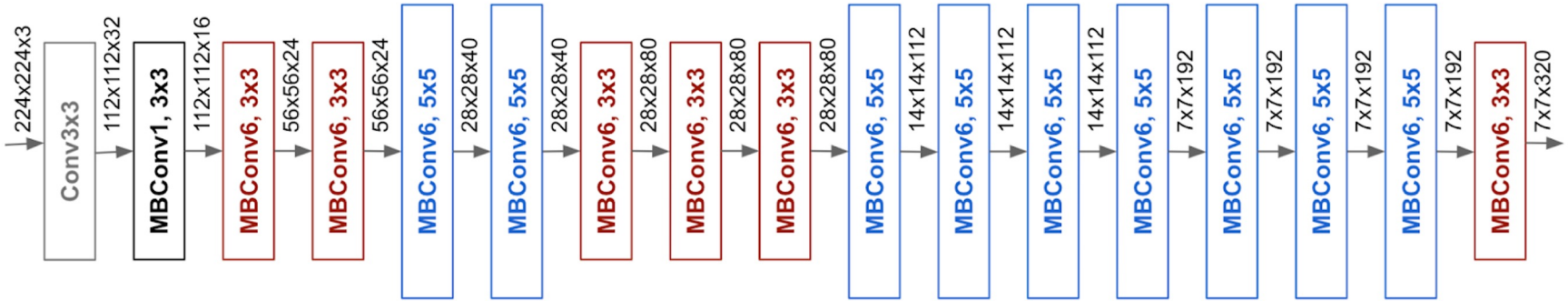
- 2 classi: cani, gatti
- Dataset bilanciato (12501 immagini per classe)

Pre trained Nets

for object classification

EfficientNet-B0

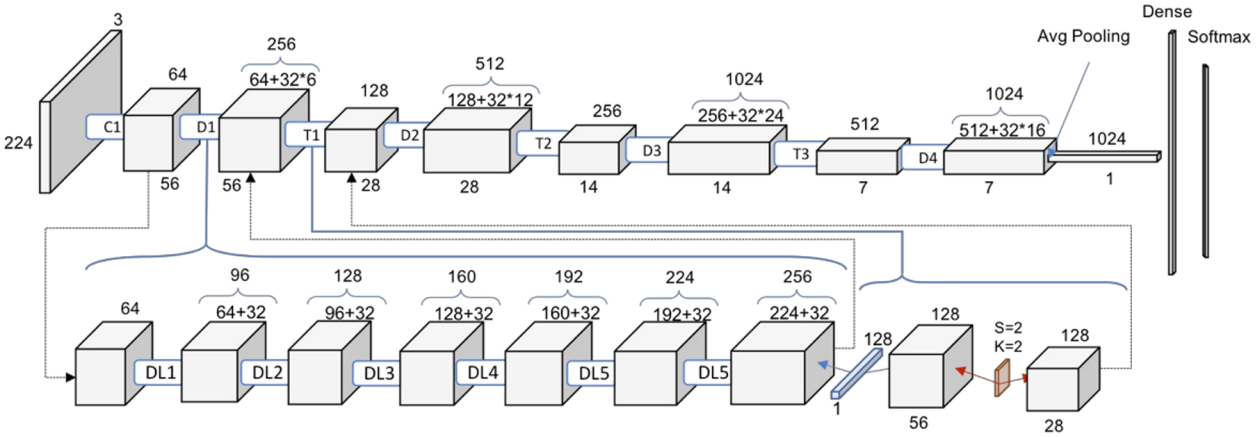
- Depth: 132 layers
- Parameters: 5.3M
- Keras function: tf.keras.applications.EfficientNetB0
- Input shape: (224, 224, 3)
- mobile inverted bottleneck convolution (MBConv), focus on parameter efficiency



EfficientNetV2-B0

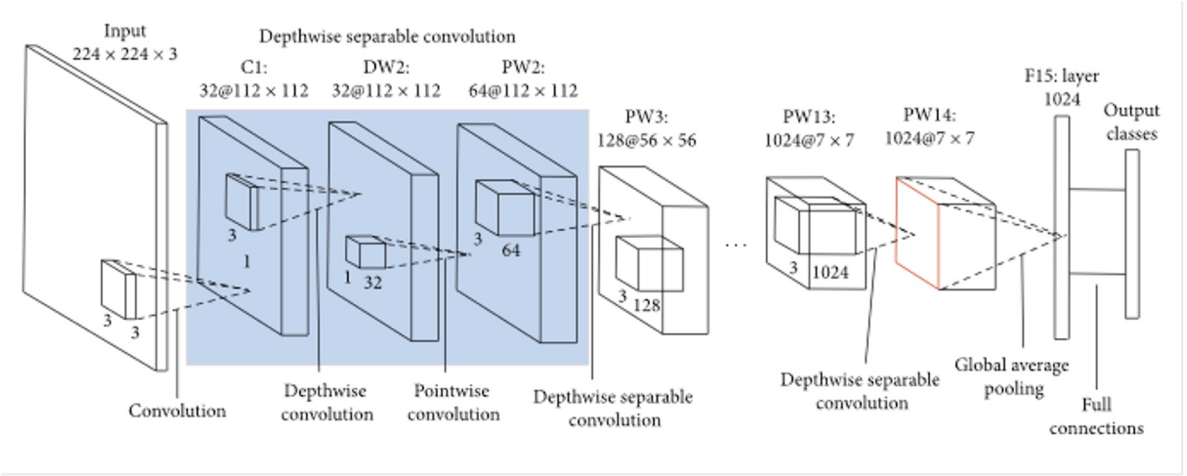
- Evolution of EfficientNet-B0
- Keras function: tf.keras.applications.EfficientNetV2B0
- Input shape: (224, 224, 3)
- Fused-MBConv, focus on parameter and computational efficiency

DenseNet121



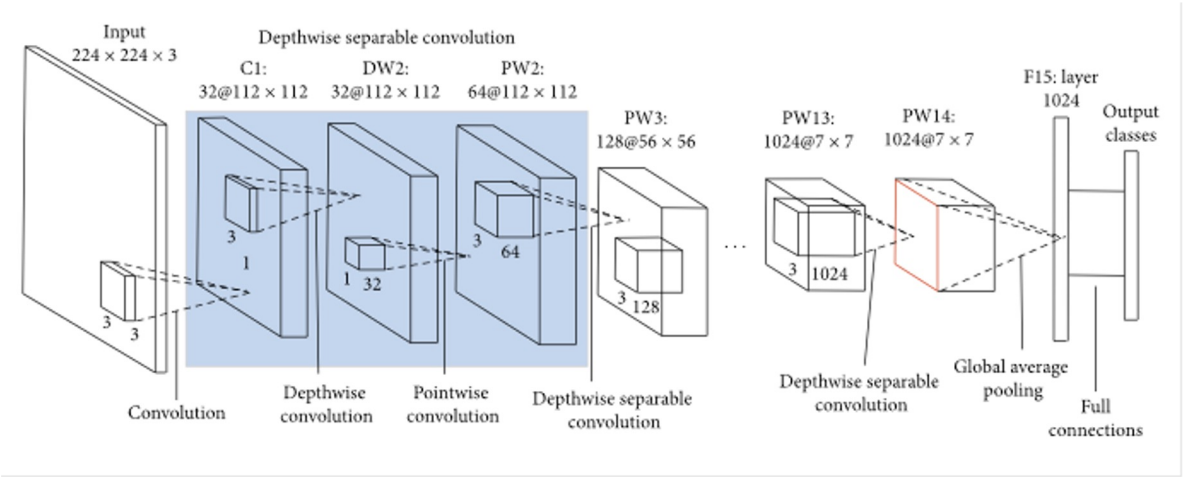
- Depth: 242 layers
- Parameters: 8.1M
- Keras function: tf.keras.applications.DenseNet121
- Input shape: (224, 224, 3)
- Densely Connected CNN

MobileNet



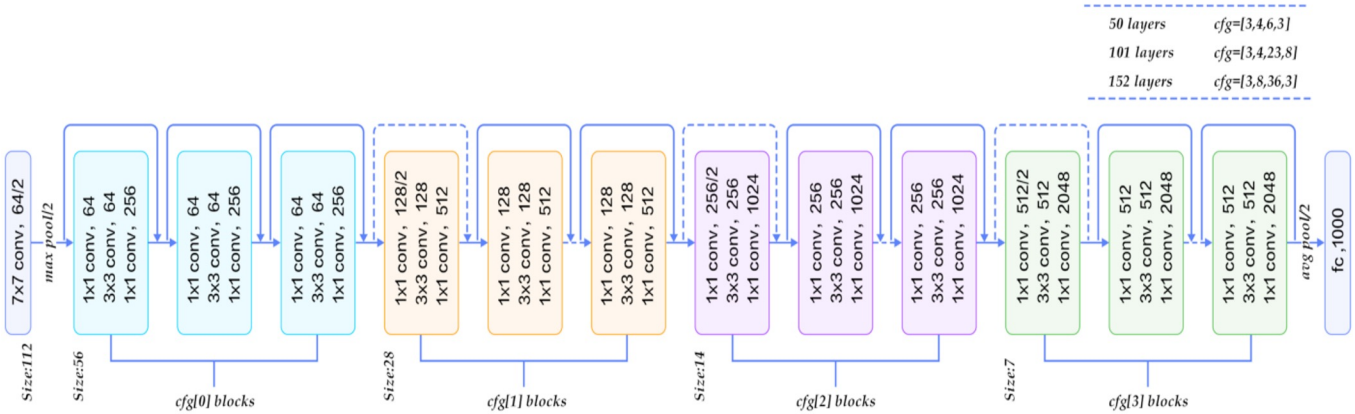
- Depth: 55 layers
- Parameters: 4.3M
- Keras function: tf.keras.applications.MobileNet
- Input shape: (224, 224, 3)
- Designed to be used in mobile applications,
- MobileNet uses depthwise separable convolutions. It significantly reduces the number of parameters

MobileNet V2



- Depth: 105 layers
- Parameters: 3.5M
- Keras function: tf.keras.applications.MobileNetV2
- Input shape: (224, 224, 3)

ResNet50

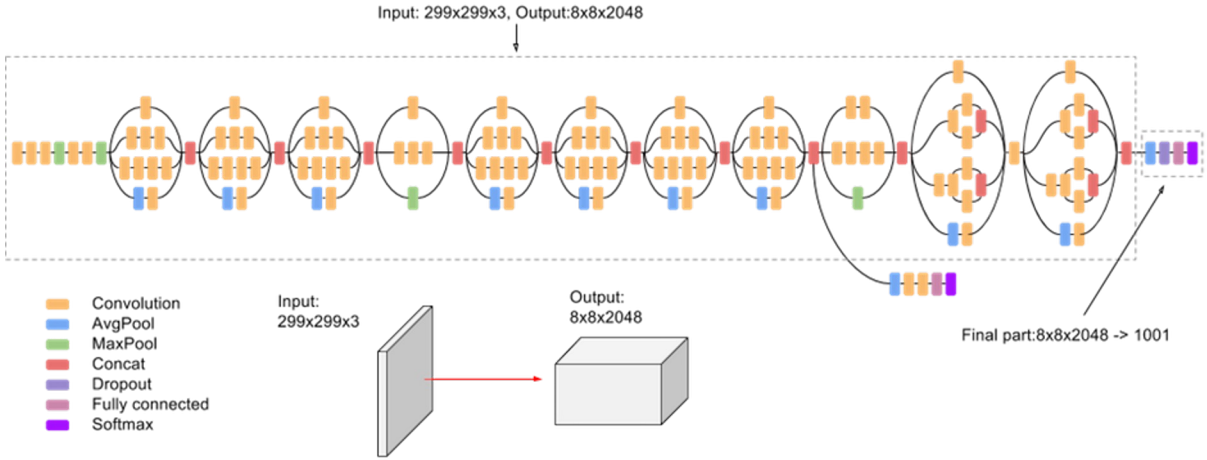


- Depth: 107 layers
- Parameters: 25.6M
- Keras function: tf.keras.applications.ResNet50
- Input shape: (224, 224, 3)
- Shortcut connections

ResNet50V2

- Depth: 103 layers
- Parameters: 25.6M
- Keras function: tf.keras.applications.ResNet50V2
- Input shape: (224, 224, 3)
- Modification in the propagation formulation of the connections between blocks

InceptionV3



- Depth: 189 layers
- Parameters: 23.9M
- Keras function: tf.keras.applications.InceptionV3
- Input shape: (299, 299, 3)

InceptionResNetV2

- Depth: 449 layers
- Parameters: 55.9M
- Keras function: tf.keras.applications.InceptionResNetV2
- Input shape: (299, 299, 3)
- Incorporates residual connections

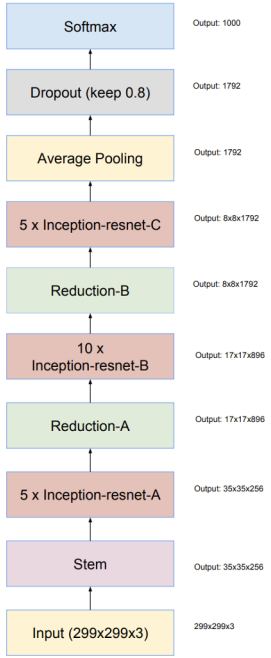
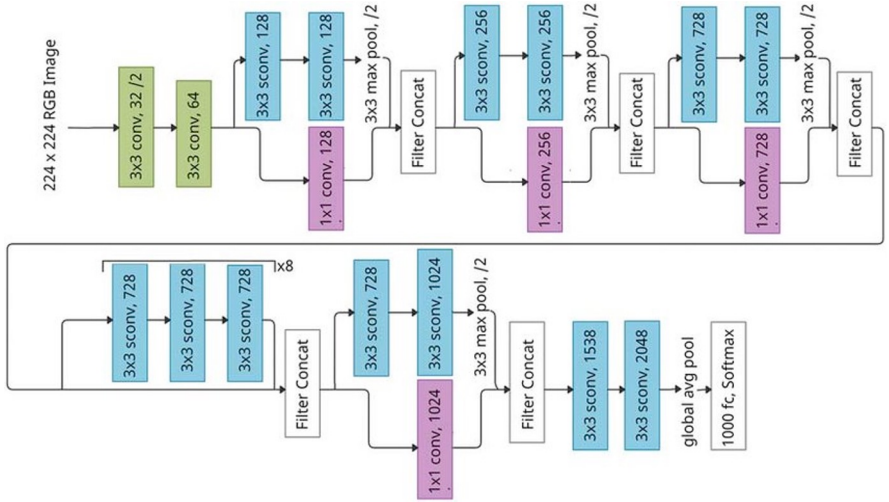


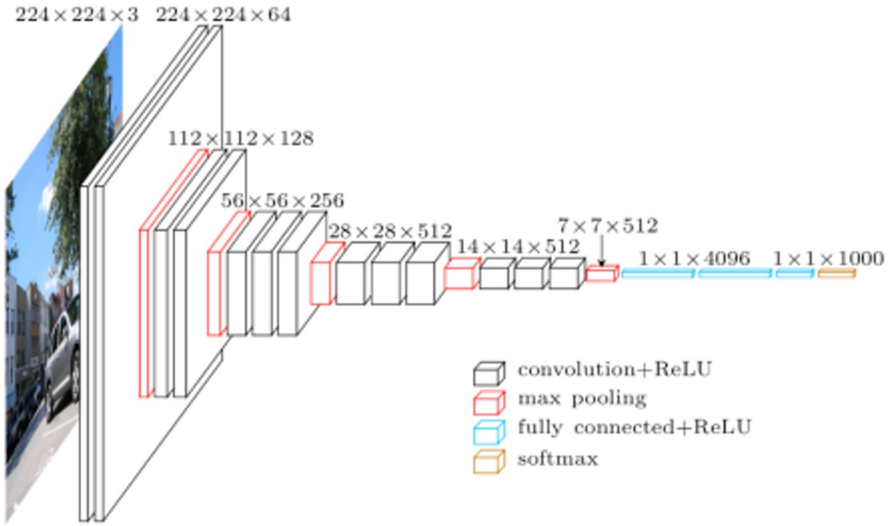
Figure 15. Schema for Inception-ResNet-v1 and Inception-ResNet-v2 networks. This schema applies to both networks but the underlying components differ. Inception-ResNet-v1 uses the blocks as described in Figures 14, 10, 7, 11, 12 and 13. Inception-ResNet-v2 uses the blocks as described in Figures 3, 16, 7,17, 18 and 19. The output sizes in the diagram refer to the activation vector tensor shapes of Inception-ResNet-v1.

Xception



- Depth: 81 layers
- Parameters: 22.9M
- Keras function: tf.keras.applications.Xception
- Input shape: (299, 299, 3)
- Inception modules have been replaced with depthwise separable convolutions

VGG16



- Depth: 16 layers
- Parameters: 138.4M
- Keras function: tf.keras.applications.VGG16
- Input shape: (224, 224, 3)

Usage

```
base_model = tf.keras.applications.EfficientNetB0(weights='imagenet', include_top=False)

x = GlobalAveragePooling2D()(base_model.output)
x = BatchNormalization()(x)
predictions = Dense(num_classes, activation='softmax')(x)

model = Model(inputs=base_model.input, outputs=predictions)

base_model.trainable = False
```

https://colab.research.google.com/drive/18xodGO4O6vsOX26lYidzN4nY63xct6zf?usp=sharing

Tabular Datasets

1. Palmer Penguins

- Predict the species of the penguins / find clusters of penguins
- Instances: **344**
- Features (# of columns): **7**
- Data type: **numerical** and **categorical**
- Missing data: **yes**
- Balanced: **no** (Adelie (152), Gentoo (124), Chinstrap (68))
- Task(s): **classification** (multi-class), **clustering**
- Info (Kaggle)

	species	island	culmen_length_mm	culmen_depth_mm	flipper_length_mm	body_mass_g	sex
0	Adelie	Torgersen	39.1	18.7	181.0	3750.0	MALE
1	Adelie	Torgersen	39.5	17.4	186.0	3800.0	FEMALE
2	Adelie	Torgersen	40.3	18.0	195.0	3250.0	FEMALE
3	Adelie	Torgersen	NaN	NaN	NaN	NaN	NaN
4	Adelie	Torgersen	36.7	19.3	193.0	3450.0	FEMALE

2. Abalone

- Predict the age* (# of Rings **OR** young (less than 10 rings)/middle age (between 10 and 20 rings)/old (between 20 and 30 rings)) of abalone from physical measurements
- Instances: **4177**
- Features (# of columns): **9**
- Data type: **numerical** and **categorical** (M/F string)
- Missing data: **No**
- Balanced: **no**
- Task(s): **classification/regression** (multi-class), **clustering**
- Info (UCI)

Sex	Length	Diameter	Height	Whole weight	Shucked weight	Viscera weight	Shell weight	Rings
M	0.455	0.365	0.095	0.5140	0.2245	0.1010	0.150	15
M	0.350	0.265	0.090	0.2255	0.0995	0.0485	0.070	7
F	0.530	0.420	0.135	0.6770	0.2565	0.1415	0.210	9
M	0.440	0.365	0.125	0.5160	0.2155	0.1140	0.155	10
I	0.330	0.255	0.080	0.2050	0.0895	0.0395	0.055	7

Sex	Length	Diameter	Height	Whole_weight	Shucked_weight	Viscera_weight	Shell_weight	Age
F	0.660	0.535	0.175	1.5175	0.7110	0.3125	0.415	MiddleAge
F	0.635	0.495	0.180	1.5960	0.6170	0.3170	0.370	MiddleAge
I	0.530	0.395	0.115	0.4750	0.2025	0.1010	0.148	Young
F	0.525	0.430	0.165	0.7170	0.2890	0.1745	0.195	MiddleAge
M	0.620	0.500	0.180	1.3915	0.7260	0.2795	0.332	MiddleAge

* da https://medium.com/ai-techsystems/abalone-age-prediction-on-cainvas-e3ac8fc27be5

3. Raisin

- Classify raisins based on morphological features extracted from images.
- Instances: **900**
- Features (# of columns): **7**
- Data type: **numerical** and **categorical** (class only)
- Missing data: **no** (missing values replaced with 0)
- Balanced: **yes** (450 per class)
- Task(s): **classification**
- Info (Kaggle)

Area	MajorAxisLength	MinorAxisLength	Eccentricity	ConvexArea	Extent	Perimeter	Class
87524	442.246011	253.291155	0.819738	90546	0.758651	1184.040	Kecimen
75166	406.690687	243.032436	0.801805	78789	0.684130	1121.786	Kecimen
90856	442.267048	266.328318	0.798354	93717	0.637613	1208.575	Kecimen
45928	286.540559	208.760042	0.684989	47336	0.699599	844.162	Kecimen
79408	352.190770	290.827533	0.564011	81463	0.792772	1073.251	Kecimen

4. Wheat Seeds Dataset

- Predict the type (A,B,C) of the seeds / find clusters of seeds
- Instances: **210**
- Features (# of columns): **8**
- Data type: **numerical** and **categorical** (class only)
- Missing data: **no**
- Balanced: **yes** (70 instances per class)
- Task(s): **classification** (multi-class), **clustering**
- Info (UCI)

Area	Perimeter	Compactness	KernelLength	KernelWidth	AsymmetryCoeff	KernelGroove	Type
14.11	14.26	0.8722	5.520	3.168	2.688	5.219	A
12.08	13.23	0.8664	5.099	2.936	1.415	4.961	A
11.19	13.05	0.8253	5.250	2.675	5.813	5.219	C
12.36	13.19	0.8923	5.076	3.042	3.220	4.605	A
16.17	15.38	0.8588	5.762	3.387	4.286	5.703	B

5a. Wine quality - red

- Predict the quality of red wines
- Instances: **1599**
- Features (# of columns): **12** (not all features are relevant ⇒ feature selection)
- Data type: **numerical**
- Missing data: **no**
- Balanced: **no**
- Task(s): **classification** (multi-class, NUMERIC!), **regression**
- Info (UCI)

fixed_acidity	volatile_acidity	citric_acid	residual_sugar	chlorides	...	density	pH	sulphates	alcohol	quality
7.5	0.580	0.56	3.1	0.153	...	0.99476	3.21	1.03	11.6	6
9.6	0.560	0.23	3.4	0.102	...	0.99960	3.30	0.65	10.1	5
10.7	0.900	0.34	6.6	0.112	...	1.00289	3.22	0.68	9.3	5
9.3	0.330	0.45	1.5	0.057	...	0.99498	3.18	0.89	11.1	7
9.1	0.765	0.04	1.6	0.078	...	0.99800	3.29	0.54	9.7	4

5b. Wine quality - white

- Predict the quality of white wines
- Instances: **4898**
- Features (# of columns): **12** (not all features are relevant ⇒ feature selection)
- Data type: **numerical**
- Missing data: **no**
- Balanced: **no**
- Task(s): **classification** (multi-class, NUMERIC!), **regression**
- Info (UCI)

fixed_acidity	volatile_acidity	citric_acid	residual_sugar	chlorides	...	density	pH	sulphates	alcohol	quality
6.4	0.33	0.30	7.2	0.041	...	0.99331	3.22	0.49	11.1	6
6.6	0.23	0.30	4.6	0.060	...	0.99142	3.23	0.49	12.2	8
6.5	0.41	0.64	11.8	0.065	...	0.99780	3.12	0.51	8.9	5
6.1	0.37	0.36	4.7	0.035	...	0.99100	3.31	0.62	12.6	6
7.5	0.33	0.28	4.9	0.042	...	0.99385	3.36	0.57	10.9	6

6. Titanic

- Predict whether Titanic passengers survived the accident or not
- Instances: **891**
- Features (# of columns): **12** (not all useful)
- Data type: **categorical** (male/female) and **numerical**
- Missing data: **yes**
- Balanced: **no** (549 no, 342 yes)
- Task(s): **classification**
- Info (Kaggle)

PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
1	No	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN	S
2	Yes	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85	C
3	Yes	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN	S
4	Yes	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123	S
5	No	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	NaN	S

7. Breast Cancer Wisconsin (Diagnostic)

- Predict the diagnosis of breast tissues (M = malignant, B = benign)
- Instances: **569**
- Features (# of columns): **32**
- Data type: **numerical**
- Missing data: **no**
- Balanced: **no** (357 benign, 212 malignant)
- Task(s): **classification**
- Info (Kaggle)

id	diagnosis	radius_mean	texture_mean	...	concavity_worst	concave points_worst	symmetry_worst	fractal_dimension_worst
842302	M	17.99	10.38	...	0.7119	0.2654	0.4601	0.11890
842517	M	20.57	17.77	...	0.2416	0.1860	0.2750	0.08902
84300903	M	19.69	21.25	...	0.4504	0.2430	0.3613	0.08758
84348301	M	11.42	20.38	...	0.6869	0.2575	0.6638	0.17300
84358402	M	20.29	14.34	...	0.4000	0.1625	0.2364	0.07678

8. Parkinson

- Predict whether a patient has the Parkinson's disease (PD) or not ((H)ealthy)
- Instances: **195**
- Features (# of columns): **24**
- Data type: **numerical**
- Missing data: **no**
- Balanced: **no** (147 PD, 48 H)
- Task(s): **classification**
- Info (UCI)

name	MDVP_Fo(Hz)	MDVP_Fhi(Hz)	MDVP_Flo(Hz)	MDVP_Jitter(%)	...	DFA	spread1	spread2	D2	PPE
phon_R01_S01_1	119.992	157.302	74.997	0.00784	...	0.815285	-4.813031	0.266482	2.301442	0.284654
phon_R01_S01_2	122.400	148.650	113.819	0.00968	...	0.819521	-4.075192	0.335590	2.486855	0.368674
phon_R01_S01_3	116.682	131.111	111.555	0.01050	...	0.825288	-4.443179	0.311173	2.342259	0.332634
phon_R01_S01_4	116.676	137.871	111.366	0.00997	...	0.819235	-4.117501	0.334147	2.405554	0.368975
phon_R01_S01_5	116.014	141.781	110.655	0.01284	...	0.823484	-3.747787	0.234513	2.332180	0.410335

9. Vertebral Column

- Classify orthopaedic patients into 3 classes (normal, disk hernia or spondilolysthesis) or 2 classes (normal or abnormal)
- Instances: **310**
- Features (# of columns): **7**
- Data type: **numerical** (class is categorical)
- Missing data: **no**
- Balanced: **no** (Normal (100), Disk Hernia (60), Spondylolisthesis (150) / Abnormal (210))
- Task(s): **classification** (binary or multi-class)
- Info (UCI)

pelvic_incidence	pelvic_tilt	lumbar_lordosis_angle	sacral_slope	pelvic_radius	degree_spondylolisthesis	class
86.47	40.30	61.14	46.17	97.40	55.75	AB
61.82	13.60	64.00	48.22	121.78	1.30	NO
33.79	3.68	25.50	30.11	128.33	-1.78	NO
60.04	14.31	58.04	45.73	105.13	30.41	AB
35.70	19.44	20.70	16.26	137.54	-0.26	AB

pelvic_incidence	pelvic_tilt	lumbar_lordosis_angle	sacral_slope	pelvic_radius	degree_spondylolisthesis	class
40.56	17.98	34.00	22.58	121.05	-1.54	DH
64.26	14.50	43.90	49.76	115.39	5.95	NO
96.66	19.46	90.21	77.20	120.67	64.08	SL
49.71	13.04	31.33	36.67	108.65	-7.83	DH
63.79	21.35	66.00	42.45	119.55	12.38	NO

Where

Datasets downloadable from

https://drive.google.com/drive/folders/1AUqQUuokZxkU9V9snxG3gMKOrlmdTUGh?usp=sharing

(only with Unife account)