

data_exploration_interim_AffectNet

February 6, 2024

1 Data exploration on AffectNet raw dataset

In this notebook are shown the main features of the dataset and the distribution of the data for the AffectNet dataset (concretely the AffectNet8Labels a reduced version from full). The generated interim dataset from the raw data will be used to perform the data exploration, as in this format it will be easier to compute the necessary operations.

First I will import the necessary libraries and load the dataset and see an example of the data and its general annotations. Afterwards, I will perform a data sanity check to avoid having issues later on. Then I will see the distribution of the data in the dataset and the distribution of the annotations in general. Finally, I will analyze it in more detail for each data split given and for each original db. Here is the main schematic of the data exploration: * 0. Interim data loading * 1. Data sanity check * 2. Data exploration over all data * 3. Data exploration over each split * 4. Conclusions

Note: For visualizations I used *blue* to denote quantities or counts and *yellow/orange* to show probabilities or normalized values.

2 0. Interim data loading

First I declare the necessary variables and import the necessary libraries.

Now we load the interim dataset:

The annotations found are for the datasplits: ['train', 'val']

Now we see some rows of the train interim dataset:

	path	cat_emot	valence	\
0	/mnt/gpid08/datasets/affectnet/raw/train_set/i...	0	-0.392238	
1	/mnt/gpid08/datasets/affectnet/raw/train_set/i...	3	0.303845	
2	/mnt/gpid08/datasets/affectnet/raw/train_set/i...	1	0.479094	
3	/mnt/gpid08/datasets/affectnet/raw/train_set/i...	1	0.787009	
4	/mnt/gpid08/datasets/affectnet/raw/train_set/i...	3	-0.015873	

	arousal
0	-0.192678
1	0.557880
2	0.077429
3	0.358637

4 0.626984

3 1. Data sanity check

Now I will perform a sanity check on the data to see if there are any issues with it. I will check that columns are equal to standard:

The columns are equal to standard in the train split

The columns are equal to standard in the val split

Now check that data types are correct:

```
Data Split: train
path          object
cat_emot      int64
valence       float64
arousal       float64
dtype: object
```

```
Data Split: val
path          object
cat_emot      int64
valence       float64
arousal       float64
dtype: object
```

And finally check that there are not null or NaN values:

```
Data Split: train
path          0
cat_emot      0
valence       0
arousal       0
dtype: int64
```

```
Data Split: val
path          0
cat_emot      0
valence       0
arousal       0
dtype: int64
```

Not null values are present, so the data sanity is correct. I can now proceed to the data exploration.

4 2. Data exploration over all data

First I will see the distribution of the data over all data splits. This way we can confirm how annotations are distributed over the whole dataset and then observe how many. I will see the

distribution of the data in the following ways: * 2.1 Continuous emotions * 2.2 Categorical emotions
* 2.3 Continuous emotions conditioned to categorical emotions * 2.4 Image size distribution

AffectNet paper sum-up for out project:

According to the AffectNet paper, all photos were queried from the web. Emotion-related keywords were combined with words related to gender, age, or ethnicity, to obtain nearly 362 strings in the English language such as “joyful girl”, “blissful Spanish man”, “furious young lady”, “astonished senior”. These keywords are then translated into five other languages: Spanish, Portuguese, German, Arabic and Farsi using native speakers and inspected the results visually. The way to select high quality-queries was done by selecting the ones that returned a high percentage of humans faces. A total of 1250 queries was done, and as many returned neutral faces, no individual query was done.

They used Google, Bing and Yahoo engines and added to the query negative terms (drawing, cartoon,...) to avoid non-human objects as much as possible. Also some photos were filtered out (this way they avoid watermarks and unnatural emotions or poses). The OpenCV library was used to detect faces in the images.

The average estimated age of the faces is 33.01 years with the standard deviation of 16.96 years. In particular, 10.85, 3.9, 30.19, 26.86, 14.46, and 13.75 percent of the faces are in age ranges [0, 10), [10, 20), [20, 30), [30, 40), [40, 50) and [50, -), respectively. MS face API (for face-keypoint generation that we are not interested in) detected forehead, mouth, and eye occlusions in 4.5, 1.08, and 0.49 percent of the images, respectively. Also, 9.63% of the faces wear glasses, 51.07 and 41.4% of the faces have eye and lip make-ups, respectively. In terms of head pose, the average estimated pitch, yaw, roll are 0.0, -0.7, and -1.19 degrees, respectively (that is the head is slightly tilted to the left).

For annotating AffectNet creators hired 12 full-time and parttime expert annotators at the University of Denver to label the database. A software application was created to annotate valence and arousal and the categorical emotion. The image below (Figure 2 in paper) shows an example of the annotation application. For each of the annotators a tutorial was given for each one of them, and three training sessions were given to each one (annotating 200 images and then analyzing then reviewing the results and feedback). Also an occlusion criterion was defined. If the person in the images wore glasses, but the eyes were visible without any shadow, it was not considered as occlusion.

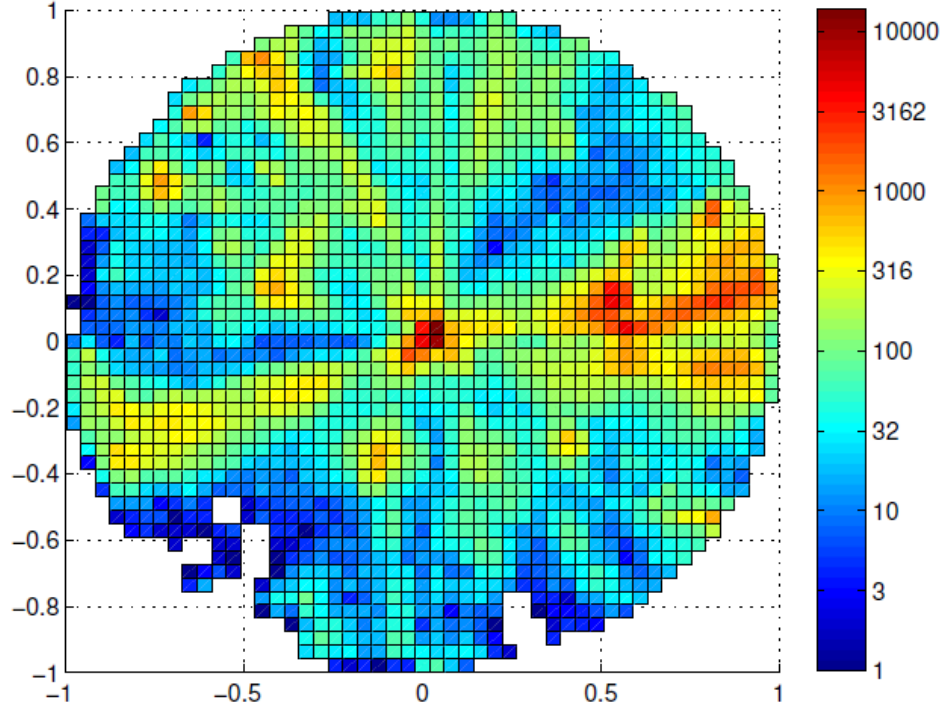


Fig. 4. Histogram (number of frames in each range/area) of valence and arousal annotations (Best viewed in color).

To model the continuous annotations, they have created a 2D Cartesian coordinate system where the x-axis and y-axis represent the valence and arousal. The annotation software did not allow the value of valence and arousal outside of the circumplex (of radius 1). This allows to convert the Cartesian coordinates to polar coordinates with r inside $0 \leq r \leq 1$ and θ inside $0 \leq \theta < 360$. Moreover, a predefined region is defined for each categorical emotion in the annotation software (e.g., for happy emotion the valence is in $(0.0, 1.0]$, and the arousal is in $[-0.2, 0.5]$). If the annotators decided to choose a value outside of the selected emotion's region, the annotation software showed a warning and it could only be there if they were sure. Also it was set a flag to be able to be reviewed by the authors. This helped to avoid mistakes in the annotation of the dimensional model.

Finally to know the measure of agreement between the annotators, 36,000 images were annotated by two annotators. The annotations were performed fully blind and independently. The results showed that the annotators agreed on 60.7% of the images.

Now I will compute the joint data distribution across all splits, this way we can analyze the distribution of the data in general.

4.1 2.1 Continuous emotions

First I will start by analyzing the distribution of the continuous emotions. In this case there are two main variables, they are arousal and valence. 1. **Valence**: The negativity/positivity of the emotion felt. It is a value between -1 and 1, where -1 is the most negative valence and 1 is the most

positive valence. 2. **Arousal**: The intensivity of the emotion felt. It is a value between -1 and 1, where -1 is the most calm and 1 is the most excited emotion.

Both draw a unit circle around 0,0 (following the AffectNet paper information), where “x” is the valence and “y” is the arousal. Now I will analyze the main statistics for the distribution of continuous emotions:

Continuous emotion:	Max	Min	Mean	Q1	Q2 (Median)	Q3
Arousal	1.0	-0.97	0.12	-0.06	0.07	0.24
Valence	1.0	-1.0	0.18	-0.19	0.19	0.65

We can observe that both of the emotions are continuous and have a range of almost [-1, 1]. The mean of both the emotions is close to 0.1 (arousal) and 0.2 (valence), which means that the dataset will have a bias to predict positive emotions with little excitement.

From the raw data exploration, in the paper we had seen that non-face had a value -2 in both arousal and valence. But as it's not present any value lower than -1 we will see later on if they are in the categorical emotions.

Now I will show the distribution of the continuous emotions in the dataset:

```
alt.HConcatChart(...)
```

We can observe how the data is distributed by the continuous emotions. We can see that the data is more concentrated in the positive emotions (valence > 0) and in the excited emotions (arousal > 0). This means that the dataset will have a bias to predict positive emotions with little excitement. Regarding the uniformity of the distribution, arousal shows a better uniformity than valence, which is more concentrated around the middle (with a huge spike around [0-0.2] interval) and more weight in the positive sector. We can also observe that there is almost no value around -1.0 and -0.8.

Now I will create a heatmap (representing a 2d histogram where color encodes the count) similar to the one that is shown in the paper to see if the results are similar. First I will create a function to create the heatmap:

Convert to pandas to proper display.

Define the max value of the plot to set the color range.

Add dotted lines to help to identify regions.

Create the plots:

```
alt.HConcatChart(...)
```

In the linear scale it is hard to observe the distribution of the data, only it can be seen a huge spike around the middle and some relevant sectors around positive/negative valences and middle arousals. For the log scale the distribution is more clear, we can observe that the data is more concentrated in the middle with some patterns from the middle to outside the circle. If we compare to the figure provided by the paper (Fig.4), the same patterns can be observed in general (good

signal as the data is distributed in the same way). In section 2.3 I will focus on analyzing the regions for each categorical emotion.

Nonetheless, if we focus our attention on the plots, we can observe little differences. More notorious around the values lower than -0.8 in arousal. There is missing rows, so we need to see how many values are missing.

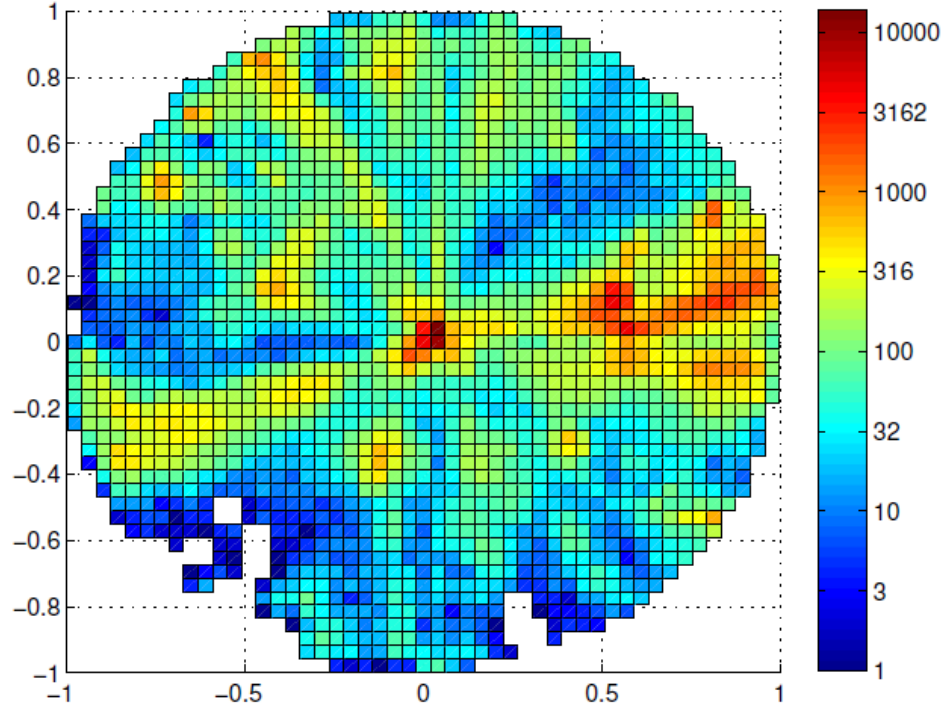


Fig. 4. Histogram (number of frames in each range/area) of valence and arousal annotations (Best viewed in color).

Now I will compare to given table by the paper (Table 14) where it is shown all the counts for each bin (of length 0.2 for each dimension). First we compute the values for the loaded dataset. We will use the same bins as in the paper and show it in a table:

+-----+-----+-----+-----+-----+-----+						
-----+-----+-----+-----+-----+-----+						
Arousal/Valence	-1.0 - -0.8	-0.8 - -0.6	-0.6 - -0.4	-0.4 - -0.2	-0.2 - 0.0	0.0 - 0.2
-----+-----+-----+-----+-----+-----+						
-----+-----+-----+-----+-----+-----+						
1.0 - 0.8	0.0	0.0	2538.0	1228.0		
3407.0	1445.0	1289.0	494.0	0.0	0.0	
0.8 - 0.6	0.0	2560.0	6059.0	3475.0		
3978.0	2127.0	2902.0	623.0	533.0	0.0	
0.6 - 0.4	468.0	6389.0	1741.0	2967.0		

1599.0		1872.0		1263.0		329.0		832.0		405.0	
	0.4 - 0.2		454.0		862.0		2835.0		4969.0		
1755.0		648.0		241.0		3859.0		8084.0		7254.0	
	0.2 - 0.0		162.0		201.0		4123.0		2753.0		
6997.0		18620.0		3683.0		37464.0		20727.0		18037.0	
	0.0 - -0.2		1942.0		765.0		2295.0		3724.0		
11496.0		13315.0		2615.0		8180.0		11755.0		7884.0	
	-0.2 - -0.4		4765.0		7944.0		3130.0		2132.0		
3386.0		1729.0		1270.0		446.0		1693.0		844.0	
	-0.4 - -0.6		441.0		495.0		150.0		217.0		
1812.0		444.0		120.0		328.0		1599.0		80.0	
	-0.6 - -0.8		0.0		13.0		21.0		30.0		
117.0		26.0		8.0		34.0		125.0		0.0	
	-0.8 - -1.0		0.0		0.0		2.0		8.0		
1.0		3.0		5.0		10.0		0.0		0.0	
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----											
-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----											

Now we will compare to the annotated oimages in the original paper (Table 14), the values are shown in the following table:

+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----											
-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----											
	Arousal/Valence		-1.0 - -0.8		-0.8 - -0.6		-0.6 - -0.4		-0.4 - -0.2		-0.2
-	0.0		0.0 - 0.2		0.2 - 0.4		0.4 - 0.6		0.6 - 0.8		0.8 - 1.0
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----											
-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----											
	1.0 - 0.8		0		0		4123		1759		
4836		1845		1672		739		0		0	
	0.8 - 0.6		0		3537		8004		4374		
5066		3379		4169		944		873		0	
	0.6 - 0.4		610		7800		2645		3571		
2042		2517		1993		467		1271		921	
	0.4 - 0.2		787		1364		3700		6344		
2804		1745		821		5241		10619		9934	
	0.2 - 0.0		395		577		5422		3675		
9024		23201		6237		42219		23281		21040	
	0.0 - -0.2		3331		1286		2971		4854		
14083		15300		4104		9998		13842		9884	
	-0.2 - -0.4		6770		9283		3884		2473		
5530		2296		3506		1824		2667		1125	
	-0.4 - -0.6		638		720		312		505		
2689		1905		1228		992		3891		957	
	-0.6 - -0.8		0		74		161		561		
706		1006		432		738		530		0	
	-0.8 - -1.0		0		0		21		674		
1021		521		60		57		0		0	
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----											
-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----											

+-----+-----+-----+-----+-----+						
+-----+-----+-----+-----+-----+						
Arousal/Valence	-1.0 - -0.8	-0.8 - -0.6	-0.6 - -0.4	-0.4 - -0.2	-0.2 - 0.0	0.0 - 0.2
+-----+-----+-----+-----+-----+						
+-----+-----+-----+-----+-----+						
1.0 - 0.8	0.0	0.0	1585.0	531.0		
1429.0	400.0	383.0	245.0	0.0	0.0	
0.8 - 0.6	0.0	977.0	1945.0	899.0		
1088.0	1252.0	1267.0	321.0	340.0	0.0	
0.6 - 0.4	142.0	1411.0	904.0	604.0		
443.0	645.0	730.0	138.0	439.0	516.0	
0.4 - 0.2	333.0	502.0	865.0	1375.0		
1049.0	1097.0	580.0	1382.0	2535.0	2680.0	
0.2 - 0.0	233.0	376.0	1299.0	922.0		
2027.0	4581.0	2554.0	4755.0	2554.0	3003.0	
0.0 - -0.2	1389.0	521.0	676.0	1130.0		
2587.0	1985.0	1489.0	1818.0	2087.0	2000.0	
-0.2 - -0.4	2005.0	1339.0	754.0	341.0		
2144.0	567.0	2236.0	1378.0	974.0	281.0	
-0.4 - -0.6	197.0	225.0	162.0	288.0		
877.0	1461.0	1108.0	664.0	2292.0	877.0	
-0.6 - -0.8	0.0	61.0	140.0	531.0		
589.0	980.0	424.0	704.0	405.0	0.0	
-0.8 - -1.0	0.0	0.0	19.0	666.0		
1020.0	518.0	55.0	47.0	0.0	0.0	
+-----+-----+-----+-----+-----+						
+-----+-----+-----+-----+-----+						

```
alt.Chart(...)
```

```
/tmp/ipykernel_819356/154834002.py:2: RuntimeWarning: invalid value encountered
in scalar divide
```

```
alt.Chart(...)
```

8

information about them.

4.2 2.2 Categorical emotions

Now I will follow my analysis with categorical emotions, these can be felt disjointly (so one face can only feel one emotion). As I have the reduced dataset, I expect only to have 8 emotions: * 0: Neutral * 1: Happy * 2: Sad * 3: Surprise * 4: Fear * 5: Disgust * 6: Anger * 7: Contempt (antipahty, disgust, or hatred)

The simplest distinction between them is that disgust is usually about physical objects (e.g., rotten food), whereas contempt is about human characteristics (e.g., incompetence). The None (“None of the eight emotions”) category is the type of expression/emotions (such as sleepy, bored, tired, seducing, confuse, shame, focused, etc.)

The annotation process was done by selecting the proper experssion category, where the intensity is not relevant. According to the paper, the majority of the returned images from the search engines were happy or neutral faces. The authors believe that this is because people tend to publish their images with positive expressions rather than negative expressions. This tells us the dataset is biased towards positive emotions.

Now I will analyze the main statistics for the distribution of categorical emotions, first I will show the distribution of the categorical emotions in the dataset:

```
alt.Chart(...)
```

As it can bee observed, the dataset is biased towards positive emotions, with a huge amount of neutral and happy faces. This means that the dataset will have a bias to predict positive emotions. The least represented emotions are contempt and disgust, which are the most negative emotions. This means that the dataset will has a bias towards positive emotions with little negativity.

Now I compare these results with the given by the paper. The paper shows the following table (table 3 in paper):

Now I compute the difference:

	cat_emot_label	difference
0	Neutral	4902
1	Happy	11283
2	Sad	3528
3	Surprise	1698
4	Fear	1313
5	Disgust	961
6	Anger	2748
7	Contempt	886
8	None	35322
9	Uncertain	13163
10	Non-Face	88895

As it is hard to compare the results, I will show it using a bar chart:

```
alt.Chart(...)
```

We can observe how **Non-Face**, **None**, **Uncertain**, **Happy** are the most affected categories in terms of quantity. Now I normalize by the size by the size of available samples in full AffectNet, this way it can be seen better how much we are losing. The first three categories were not mentioned in the notes of AffectNet8Labels, so we will see with the normalization if all are lost or not.

	cat_emot_label	percentage
0	Neutral	0.061064
1	Happy	0.077176
2	Sad	0.119646
3	Surprise	0.104249
4	Fear	0.160298
5	Disgust	0.182561
6	Anger	0.097689
7	Contempt	0.172541
8	None	1.000000
9	Uncertain	1.000000
10	Non-Face	1.000000

I display it using a bar chart to get a better sense of the loss:

```
alt.Chart(...)
```

We can observe how **None**, **Uncertain**, **Non-Face** have no representation in the given dataset. For the other categories we have around a loss of 15% of the data, which is not that much comparing the missing the other labels (maybe some data cleaning step was done for the AffectNet reduced version). But considering we have already few data for the negative emotions, this will have an impact on the model performance.

4.3 2.3 Continuous emotions conditioned to categorical emotions

Now I will focus my analysis on seeing the distribution of the continuous distribution conditioned to a emotion category. This way we can see in which sectors each categorical emotion is used to be. The scale used for all of them is the same as the one used before (the logarithmic).

The maximum number of records in a bin is: 14617

The logarithmic scale is: [1 4 24 120 597 2956 14617]

4.3.1 Neutral

```
alt.LayerChart(...)
```

It can be seen that **Neutral** is centered in the middle of the unit circle, with a huge spike around the middle. All samples draws almost a half unit circle, and some relevant weight (apart from the huge spike) is centered around the circle. There are also some samples outside this half unit circle, but they are not that relevant.

4.3.2 Happy

```
alt.LayerChart(...)
```

For the categorical emotion **Happy**, its samples are distributed almost all on the positive values of the valence variable. For the arousal variable is more uniformly distributed, but it has more weight towards values around 0 (positive and negative). There are also some samples outside the half of the unit circle, but they are not that relevant.

4.3.3 Sad

```
alt.LayerChart(...)
```

Sad is mostly distributed in the negative values of valence and arousal (even though it has some representation on the positive values of arousal). It does not show a clear half circle as happy. It does not show many outliers outside the mentioned intervals.

4.3.4 Surprise

```
alt.LayerChart(...)
```

Suprise is mainly located in the positive values of arousal (upper than 0.25), and it is more weighted distributed in the positive values of the valence variable. It shows some outliers in the negative values of arousal and valence (showing that in some special cases surprise can be a negative emotion).

4.3.5 Fear

```
alt.LayerChart(...)
```

Fear is located in a section in the upper left of the unit circle. So it is mainly located in the negative values of valence and positive value of arousal (bigger than 0.25 in valence and between 0.1 and -0.3 in arousal). In this case it shows some outliers from the main weight distribution, for example the cluster around -0.25 to -0.5 in arousal and valence.

4.3.6 Disgust

```
alt.LayerChart(...)
```

Disgust is locaated in the negative values of valence and positive values of arousal. The distribution it is not as the other emotions as it is a bit more scattered. It shows some outliers in the negative values of arousal.

4.3.7 Anger

```
alt.LayerChart(...)
```

Anger is distributed in the negative values of valence and positive values of arousal. It draws a quarter of the unit circle, and it shows some outliers in the negative values of arousal.

4.3.8 Contempt

```
alt.LayerChart(...)
```

Contempt is located in the negative values of valence and positive values of arousal. Its' distribution is not as clear as the other emotions as it is a bit more scattered (with a kind of more weighted line from (-0.75, 0.75) to (-0.25, 0.25) (valence,arousal). It shows some outliers following the line mentioned.

4.3.9 All emotions:

```
alt.VConcatChart(...)
```

We can observe that the unit circle is not disjoint in the categorical emotion, as there are some that are located in the same region. For example, and fear, disgust, anger and contempt distributions are located in the same region. This means that we cannot tell the categorical emotion easilly by knowing only the arousal, valence values.

4.4 2.4 Image size distribution

Now I will analyze the distribution of the image sizes in the dataset. This way we can see the distribution of the image sizes and see if there are any issues with the data. In the notes of the AffectNet8Labels dataset it is mentioned that the images are resized to 256x256, so I expect to see a huge spike around this value.

In order to make the computations faster, I will analyze 2500 photos per split and show the results. This way we can see the distribution of the image sizes in the dataset (I expect that this sample is representative of the whole dataset).

Now I will show the height and width distribution of the images in the dataset:

```
alt.HConcatChart(...)
```

We can see that all sampled photos have a size of 256x256, which is the expected value. This means that the dataset is consistent in terms of image sizes.

```
alt.Chart(...)
```

All images will have an aspect ratio of 1:1, which is the expected value. This is not realistic as the aspect ratio of the human face is not 1:1 (and the face detector is not set up to predict always this square box), but it is the expected value as the images are resized to 256x256. During the following steps we will see if this needs to be taken to account.

5 3. Data exploration over each split

I will observe how the data is distributed over each split, to see if the data distributions of them are different. I will analyze the distribution of the data in the following ways: * 3.1. Continuous emotions * 3.2. Categorical emotions

First I will analyze the amount of annotations per split:

```
Number of entries in train:    287651 (98.63% of total)
Number of entries in val:     3999 (1.37% of total)
```

5.1 3.1. Continuous emotions

Now I will analyze the distribution of the continuous emotions in each split. I will do the following:

- * 3.1.1. Visual comparison
- * 3.1.2. Normalized difference between splits
- * 3.1.3. KL divergence between splits

3.1.1. Visual comparison

I will start concatenating the resulting visualizations for the `train` and `val` data splits. I will use the same heatmap as before to show the distribution of the data. First I will compute the values for the splits:

Now I convert it to pandas dataframe.

Remember that the scale used for the heatmap is the same as the one used before (the logarithmic).

The maximum number of records in a bin is: 14617

The logarithmic scale is: [1 4 24 120 597 2956 14617]

I add the dotted lines to help to identify regions.

Show the visualizations:

```
alt.HConcatChart(...)
```

We can see that in general they have the same distribution. The main difference is that the validation split is much more scattered than the training split (it is expected, as there are less entries and the general distribution of data should be preserved).

5.1.1 3.1.2. Normalized difference between splits

Now we normalize the data to the size of the data split and subtract the two (train-val) to see the difference. This way we can see if a section in validation is over/under represented (negative/positive difference correspondingly).

```
alt.LayerChart(...)
```

The whiter zones of the image show under representation and the darker zones show over representation from the validation set with respect to the train set. We can see that the data is distributed in a similar way in both splits, but there are some differences. For example, the validation split has more data in the negative values of valence and arousal, and less data in the positive values of valence and arousal. This means that the validation split has a bias towards negative emotions with little excitement in comparison to the train split. Also in the middle of the circle there is not the same peak as in the train split.

This might be caused due to the normalization of the data on categorical emotions to have the same number of entries for each category. This way the data is not distributed in the same way as the train split but this way we ensure that we test the model on all categories equally.

5.1.2 3.1.3. KL divergence between splits

Now I will compute the KL divergence. The Kullback-Leibler (KL) divergence is a measure of how one probability distribution differs from another (it's not symmetric). In the context of discrete probability distributions, such as histograms it is defined by the following formula:

$D_{KL}(P||Q) = \sum_{x \in \mathcal{X}} P(x) \cdot \log\left(\frac{P(x)}{Q(x)}\right)$, where \mathcal{X} is the sample space

Note: Even though the real distribution is continuous, I will use the histogram to approximate the continuous distribution and compute the KL divergence. This way I can compare the distributions of the data more easily.

Note 2: Relative entropy is only defined in this way if, for all x , $Q(x) = 0$ implies $P(x) = 0$ (absolute continuity).

First we will consider the divergence for $P(x) = \text{val}$ from $Q(x) = \text{train}$ (val from train). The KL divergence of P from Q is the expected excess surprise from using Q as a model when the actual distribution is P .

In the simple case, a relative entropy of 0 indicates that the two distributions in question have identical quantities of information. And the highest one is infinite.

The KL divergence $D(\text{val}||\text{train})$ is: 0.74380

We can see that if we use train as a model (the reality of the data), when the actual distribution is P (val), we have a KL divergence of 0.74. This means that they are similar distributions with the weight distributed in general with the same way.

Now we will consider the divergence for $P(x) = \text{train}$ from $Q(x) = \text{val}$:

The KL divergence $D(\text{train}||\text{val})$ is: 0.49431

In this case the KL divergence is 0.49, using the val split as a model when the actual distribution is train. This is expected by definition, as the KL divergence is not symmetric. The divergence is lower when using val as a model. As $D(\text{train} || \text{val})$ is smaller than $D(\text{val} || \text{train})$, it suggests that more information is being transferred from train to val than in the reverse direction (this additional information is expected as train split should be more representative). In other words, **val** captures more of the essential features of **train** than vice versa.

In both cases we have a relatively low KL divergence, which means that the distributions are similar in the relevant zones. This means that the data is distributed similarly in both splits for the continuous emotions, having more information from train to val than in the reverse direction (that is expected, as train should be more representative and the model should learn more from it).

5.2 3.2. Categorical emotions

Now I will analyze the distribution of the categorical emotions in each split. I will use the same bar chart as before to show the distribution of the data. First I will compute the values for the splits:

```
alt.HConcatChart(...)
```

As the scales are difficult to compare, we will normalize by the size of the dataset and show the results in a bar chart.

```
alt.HConcatChart(...)
```

We can observe that the distribution of data is different. As shown in the paper, the validation split has a uniform distribution of categorical emotions (this way the model performance can be tested on all emotions equally), while the training split has a bias towards positive emotions due to the

results of queries. This means that the model will have a bias towards positive emotions with little negativity while training, but at test time it will be tested on all emotions equally.

Now I will compute the KL divergence for the categorical emotions. First I will consider the divergence for $P(x) = \text{val}$ from $Q(x) = \text{train}$ (val from train):

The KL divergence $D(\text{val}||\text{train})$ is: 0.72895

Now I will compute the KL divergence for the categorical emotions. First I will consider the divergence for $P(x) = \text{train}$ from $Q(x) = \text{val}$ (train from val):

The KL divergence $D(\text{train}||\text{val})$ is: 0.60107

We have similar behaviour as before. The divergence is lower when using val as a model. As $D(\text{train} || \text{val})$ is smaller than $D(\text{val} || \text{train})$, it suggests that more information is being transferred from train to val than in the reverse direction (this additional information is expected as train split should be more representative). In other words, val captures more of the essential features of train than vice versa.

6 4. Conclusions

The AffectNet dataset contains a diverse set of both continuous and categorical emotions, providing a rich resource for emotion recognition tasks. These are the conclusions of the data exploration:

6.1 Preprocessing data:

- There is missing data with respect to the original paper, as obtained dataset is **Affect-Net8Labels**. The labels **None**, **Uncertain**, **Non-Face** are not present in the dataset. This means that the dataset is not as complete as the original AffectNet dataset. Also the under part of the unit circle for categorical emotions is not as represented as in the original dataset.
- Consider to use the cartesian or polar coordinates for the continuous emotions to see the distribution of the data in a better way.
- All recieved images from AffectNet are resized to 256x256.
- Techniques like data augmentation or oversampling might be necessary to address class imbalance.
- There is no test split. So I will use the validation split as the test (evaluating the model's performance on all emotions equally), and create a validation split from training using random sampling (as if we use satrified sampling per categorical emotions, there will be few negative samples to learn from).

6.2 Training model:

- It must be taken into account the bias towards positive emotions in the training split, as there is an underrepresentation of negative emotions.
- Many performance functions will need to be used to monitor and analyze the model's performance on different emotions to ensure it is learning effectively.
- It is important to consider the aspect ratio of the images (1:1), it is not the natural bounding box of a face.

These are some general points to consider, but further on I will specify which techniques will be applied.

7 References

- A. Mollahosseini; B. Hasani; M. H. Mahoor, “AffectNet: A Database for Facial Expression, Valence, and Arousal Computing in the Wild,” in IEEE Transactions on Affective Computing, 2017. [AffectNet paper](#)