# 情感计算（*Sentiment Analysis*）资料整理

1. 情感计算概念的来源：1997 年，*MIT* 媒体实验室提出了情感计算（*Affective Computing*）的概念。情感计算旨在通过赋予计算机识别、理解和表达人的情感的能力，使得计算机具有更高的智能。

2. 情绪分类的概念：是指区分或者对比一种情绪与另一种情绪的方法，目前在情绪研究（*emotion research*）与情感科学（*affective science*）是具有争议的问题。有两个讨论情绪分类的基本观点，也即下文中即将涉及的两大情感分类体系：***Dimensional Emotion*** 和 ***Discrete Emotion***，译为维度情感和离散情感。

这两种分类体系可以简单理解为，前者是将情感分作不同衡量维度，将每一种情感映射到这一空间当中，通过数值确定唯一情感，后者则是通过对情感进行分类，将其作为一个个独立标签相互之间没有关联。这就算铺垫完成，下面将直接开始介绍这两种不同的情感分类体系。

下面的如果直接看不懂，可以翻到参考文献前面的图表，整理的应该还算好理解（吧）。

## (一) *Discrete Emotion*

*The discrete emotion model, also called as categorical emotion model, defines emotions into limited categories. Two widely used discrete emotion models are Ekman's six basic emotions and Plutchik's emotional wheel model.*

也就是目前关于离散情感分类存在两种主流的方式，分别是 *Ekman* 的六种基本情感(*basic emotion*)以及 *Plutchik* 的情感轮子模型(*emotional wheel model*)，分别展示在 *Figure 1* 中：

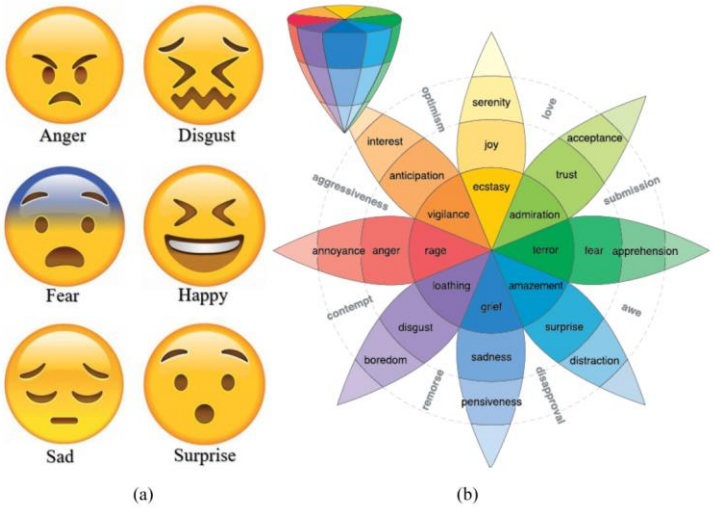



***Figure 1*: 两种 *Discrete Emotion Model***

*Ekman* 提出的六种情感主要遵循这么几个原则：首先这六种情感都是源自于人类的直觉；其次，当人们面临同样情况时，会产生同样的情感；再其次，当在同样的语义/情境中时，人们会表达出同样的情感；最后，对所有人而言，这些基本情感对所有人都应该有同样的表达模式（*pattern of expression*）。这个轮子能被提出来的前提条件是：人类情感是可以跨越种族和文化而被共享的。

*Plutchik* 的轮子模型涉及八种基本情绪(即喜悦、信任、恐惧、惊讶、悲伤、期待、愤怒和厌恶)以及这些情绪之间的相互关系。例如，快乐和悲伤是对立的，期待很容易发展成警惕。轮子模型也被称为成分模型（*componential model*），其中较强的情绪占据内测，而较弱的情绪占据外侧，这一位置的排布取决于它们的相对强度水平。这些离散的情绪通常可以分为三种极性(积极，消极和中性)。

## (二) *Dimensional Emotion*

前面提到的离散情感模型的弊端在于，不同情感之间没有类似于数值的可连续性，不同标签间的差异和连续性就无法更好的解释和计算。在维度情感模型中，最被广泛使用的是 *PAD* 模型，也即每种情感被映射到 *PAD* 的三维空间中。

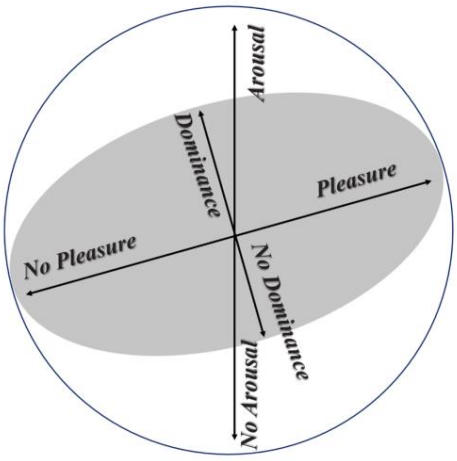



***Figure 2*: PAD Dimensional Emotion**

其中，这三个维度代表的含义如下：

*1) Pleasure (Valence) dimension, representing the magnitude of human joy from distress extreme to ecstasies;*

*2) Arousal (Activation) dimension, measuring physiological activity and psychological alertness level;*

*3) Dominance (Attention) dimension, expressing the feeling of influencing the surrounding environment and other people, or of being influenced by the surrounding environment and others.*

由于在这一模型中，*P* 和 *A* 两个维度就足以表示大多数情感，所以后面又有人提出

了一个基于 *Valence-Arousal* 的 *circumplex* 模型，很显然，*Figure3* 中的这个圆圈模型中只有 *valence* 和 *arousal* 两个维度：

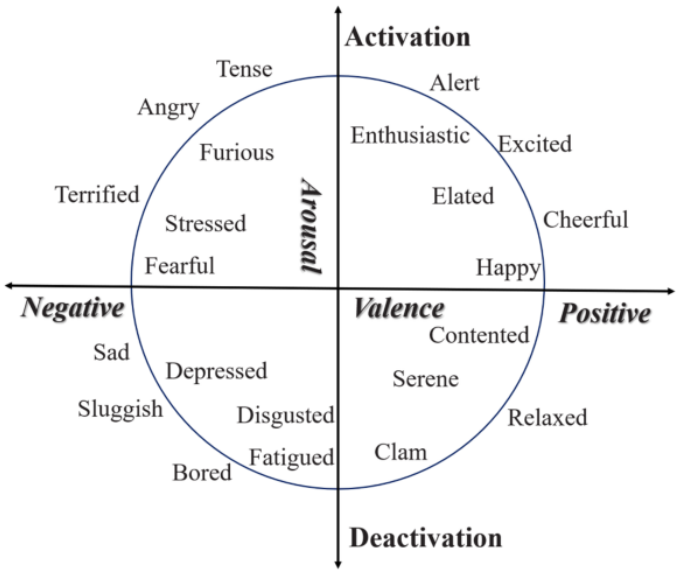



***Figure 3: Valence-Arousal Emotion Model***

*The first quadrant, activation arousal with positive valence, shows the feelings associated with happy emotions; And the third quadrant, with low arousal and negative valence, is associated with sad emotions. The second quadrant shows angry emotions within high arousal and negative valence; And the fourth quadrant shows calm emotion within low arousal and positive valence.*

在上述圆圈模型中，不同象限的含义不同。其中横轴 *valence* 直译为“效价”，可以理解为一种情感它的正负面属性，如果该情感属于正面，那么它就为“*positive valence*”，反之则为“*negative valence*”；*arousal* 直译为“唤醒”，理解为一种情感本身所带有的沉浸属性，如果一种情感使人投入状态更高，那么它就具有更高的 *arousal*，也就是让人投入的程度更深，即 *high arousal*。

在第一象限为“积极效价高唤醒”，代表性情感都具有一些正面且投入程度较高的特征；二象限为“消极低唤醒”，这一象限中的情感特征为负面但投入程度较低；三象限为“消极低唤醒”，情感负面投入程度也较低；四象限为积极低唤醒，情感正面，但让人沉浸的程度也不会很高。至此，基于 *valence-arousal* 的二维圆圈模型阐述完毕。

除了上面提到的 *PAD* 和 *valence-arousal* 模型，*dimensional emotion model* 还有其他的建模方式，如果能整理完的话，可能会放在最后再讲，接下来讲一讲不同模态间的情感计算方式。

## (三) *Uni-Modal Affective Computing/Analysis*

这一部分看大标题就知道，肯定是分两个部分，一个是 *affective recognition*，另一个是 *affective analysis*，前者涉及更多的是单模态（*unimodal*）的情感分析或识别，例如文本语音图片，乃至于生理心理上的数据分析，后者则涉及到多模态（*multimodal*）的情感分析，包括但不限于上述各个模态间的交互作用，多模态的内容会放到下一个 *chapter* 讲，如果我能整理完的话。

## 一、 文本情感识别（*Textual sentiment analysis*）

### 1. 基于机器学习的文本情感识别（*ML-based TSA*）

#### a） 基于知识的 *TSA (Knowledge-Based TSA*）

这一方法通常依赖词典或者语言的规则，不同的词典具有不同的词包和不同的语言词性注释，这一方法可以基于词典将一个单词划分为正负两极，但如果失去了词典的定义，即某种语言法则后，表现并不好。由于知识本身的局限性，基于知识的模型仅限于理解那些典型的和严格定义的概念。在这一方法的基础上，后续有人不断对其做出改进，包括考虑不同领域的词典扩充；运用计算智能(*computational intelligence*)、语言特征、常识的计算方式等。

#### b） 基于统计的 *TSA (Statistical-Based TSA*）

这一方法更多地依赖于带注释的数据集，通过使用先验统计或后验概率来训练基于 *ML* 的分类器。与基于词典的方法相比，基于统计的 *TSA* 方法更适于情感分析，因为后者可处理更多的数据。

这一方法中，使用最多的就是 *SVM*（支持向量机）和 *NB*（朴素贝叶斯），前者通过构建基于语义的特征空间，再利用 *SVM* 对其进行分类；后者则是假定不同数据集间的特征相互独立，随后以此来过滤不支持这一比较假定的句子[1]。

#### c） 基于混合方法的 *TSA (Hybrid-based TSA*）

字面意思就是结合上面两种方法，优点肯定是它俩的优点，关于这一方法有一些文献：*Le*[2]等人提出了一种新的混合方法，将单词情感得分计算、文本预处理、情感特征生成和基于 *ML* 的情感分析分类器结合在一起，这一方法比纯基于词典在 *Amazon*、*IMDb* 和 *Yelp* 有更好的表现；*Li*[3]等人弄了一套基于 *ML* 的词典方法，具体包括 *SVM* 和 *NB*，这一方法对处理文本的极性更加有效（*posi-nega* 分类）。

## 2. 基于深度学习的文本情感识别（*DL-based TSA*）

### *a)* 使用深度卷积网络的 *TSA (Deep ConvNet-Based TSA)*

这一方法已经用于不同级别的文本情感分析，文档级、句子级、单词级，通过使用不同级别的过滤器来提取文本特征。

*Yin et al. [4] proposed a framework of sentence-level sentiment classification based on the semantic lexical-augmented CNN (SCNN) model, which makes full use of word information. Conneau et al. [5] applied a very deep CNN (VDCNN), which learns the hierarchical representations of the document and long-range dependencies to text processing. To establish long-range dependencies in documents, Johnson and Zhang [6] proposed a word-level deep pyramid CNN (DPCNN) model, which stacked alternately the convolutional layer and the max-pooling down-sampling layer to form a pyramid to reduce computing complexity. The DPCNN with 15 weighted layers outperformed the previous best models on six benchmark databases for sentiment classification and topic categorization. For aspect-level sentiment analysis, Huang and Carley also [7] proposed a novel aspect-specific CNN by combining parameterized filters and parametrized gates.*

### b) 使用循环神经网络的 *TSA (RNN-Based TSA)*

循环神经网络适用于处理长序列数据，注意力机制有助于根据加权表示对给定输入序列的相关部分进行优先排序，且计算成本低。在 *TSA* 的基于注意力机制的 *LSTM* 范式中，*LSTM* 帮助构建文档表示，然后基于注意力的深度记忆层计算每个文档的评级。

*For document-level sentiment classification, Dou [8] proposed a deep memory network combining LSTM on account of the influence of users who express the sentiment and the products that are evaluated. Chen et al. [9] designed a hierarchical LSTM with an attention mechanism to generate sentence and document representations, which incorporates global user and product information to prioritize the most contributing items. Considering the irrationality of encoding user in- formation and product information as one representation, Wu et al. [10] designed an attention LSTM-based model, which executed hierarchical user attention and product attention (HUAPA) to realize sentiment classification.*

*For multi-task classification (e.g., aspect category and sentiment polarity detection), J et al. [11] proposed convolutional stacked Bi-LSTM with a multiplicative attention network concerning global-local information. In contrast, to fully exploit contextual affective knowledge in aspect-level TSA, Liang et al. [12] proposed GCN-based SenticNet to enhance graph-based dependencies of sentences. Specifically, LSTM*

*layers were employed to learn contextual representations, and GCN layers were built to capture the relationships between contextual words in specific aspects.*

### c） 使用 *DCNN-RNN* 的 *TSA (Conv-RNN- Based TSA)*

*As the CNN is proficient in extracting local features and the BiLSTM is skilled in a long sequence, Li et al. [13] combined CNN and BiLSTM in a parallel manner to extract both types of features, improving the performance of sentiment analysis. To decrease the training time and complexity of LSTM and attention mechanism for predicting the sentiment polarity, Xue et al. [14] designed a gated convolutional network with aspect embedding (GCAE) for aspect-category sentiment analysis (ACSA) and aspect-term sentiment analysis (ATSA). The GCAE uses two parallel CNNs, which output results combined with the gated unit and extended with the third CNN, extracting contextual information of aspect terms. To distinguish the importance of different features, Basiri et al. [15] proposed an attention-based CNN-RNN deep model (ABCDM), which utilized bidirectional LSTM and GRU layers to capture temporal contexts and apply the attention operations on the discriminative embeddings of outputs generated by two RNN-based networks. In addition, CNNs were employed for feature enhancement (e.g., feature dimensionality reduction and position-invariant feature extraction). All the above works focused on detecting sentiment or emotion, but it is important to predict the intensity or degree of one sentiment in the description of human intimate emotion. To address the problem, Akhtar et al. [16] proposed a stacked ensemble method by using an MLP to ensemble the outputs of the CNN, LSTM, GUR, and SVR.*

### *d)* 使用深度对抗学习的 *TSA (Adversarial -Based TSA)*

可以参考使用这一方法执行文本分类的文献[17]

## 二、 音频情感识别（*Audio Emotion Recognition*）

音频情感识别又称 *SER*(*Speech Emotion Recognition*)，是通过处理和理解语音信号来检测音频内情感，这一方法同样分为了基于 *DL* 和基于 *ML*，传统的基于 *ML* 的方法专注于提取声学特征并选择适当的分类器；*DL* 方法则通过构建端到端的 *CNN* 结构来预测情绪。

### 1. 基于机器学习的音频情感分析 *(ML-Based SER)*

基于机器学习的音频情感识别包括两个关键步骤：情感语音的强特征表示学习

（*strong features representation learning for emotional speech*）和最终情感预测的适当分类。

不同种类的声学特征可以通过特征融合得到混合特征，进而实现稳健的 *SER*，音频的质量特征（*voice-quality features*）在某些情况下比语音特征（*prosodic features*）和频谱特征（*spectral features*）在 *SER* 中更为重要。语音中关键特征的提取可以使用 *OpenSMILE* 工具。

a) **基于声学特征的 *SER*（*Acoustic-feature based SER*）**

*Prosodic features (e.g. intonation and rhythm) have been discovered to convey the most distinctive properties of emotional content for SER. The prosodic features consist of fundamental frequency (rhythmical and tonal characteristics) [18], energy (volume or the intensity), and duration (the total of time to build vowels, words and similar constructs). Voice quality is determined by the physical properties of the vocal tract such as jitter, shimmer, and harmonics to noise ratio.*

b) **基于机器学习分类的 *SER (ML-Based Classifier SER)***

*Different from HMM and GMM, SVM maps the emotion vector to a higher dimensional space by using a kernel function and establishes the maximum interval hyperplane in the high-dimensional space for optimal classification.*

## 2. 基于深度学习的音频情感分析（*DL-Based SER*）

基于 *DL* 的 *SER* 系统可以理解和检测情感语音的语境和特征，而无需设计专门的特征提取器。带有自动编码器的 *CNN* 被认为是基于 *DL* 的 *SER* 的常用技术。*RNN* 及其变体（如 *BiLSTM*）被广泛用于捕捉时间信息。用于 *SER* 的混合深度学习包括 *ConvNets* 和 *RNNs* 以及注意力机制。针对数据量有限和数据库质量不高的问题，对抗学习可通过增强训练数据和消除扰动，用于基于 *DL* 的 *SER*。

***a)* 使用卷积网络学习的 *SER (ConvNet-Based SER)***

***b)* 使用循环网络学习的 *SER (RNN-Based SER)***

***c)* 使用 *ConvNet-RNN* 的 *SER (Conv-RNN SER)***

***d)* 使用对抗学习的 *SER (Adversarial -Based SER)***

## 三、 视觉情感识别（*Visual Emotion Recognition*）

视觉情感识别可以简单概括为两类：面部情感识别（*facial expression recognition*，*FER*）和肢体情感识别（*emotional body gesture recognition*，*EBGR*）。论文中主要分为了这两种方法，在多模态部分我可能还会补充一下我自己用到的一个图片整体情感的识别方法（如果顺利的话）。

*FER* 是使用包含面部情感线索的图像或视频来实现的。根据使用静态图像还是动态视频来表示面部表情，*FER* 系统可以分为基于静态的 *FER* 和基于动态的 *FER*。在面部表情持续时间和强度方面，*FER* 又可分为宏观 *FER* 和微观 *FER*(或 *FMER*)。根据面部图像的维度，宏观 *FER* 可以进一步分为 2*D FER* 和 3*D*/4*D FER*。由于面部图像或视频受到各种背景，照明和头部姿势的影响，因此使用预处理技术(例如人脸对齐、人脸归一化和姿态归一化等)，对人脸区域的语义信息进行对齐和归一化。

## 1. 基于机器学习的 *FER*

### *a)* 基于几何学的 *FER (Geometry-Based FER)*

*Ghimire and Lee*[19]等人利用面部的关键位置点用来表示几何学位置和角度，设计了一套在面部序列数据中自动识别表情的东西；*Sujono and Gunawan*[20]使用（*Kinect motion sensor*）这一传感器来识别面部关键区域，同样应该也是运用几何学原理来预测面部表情。

*Sujono and Gunawan used the Kinect motion sensor to detect the face region based on depth information and active shape model (AAM). The change of key features in AAM and a fuzzy logic model is utilized to recognize facial expression based on prior knowledge derived from the facial action coding system (FACS).*

### b) 基于外观的 *FER (Appearance-Based FER)*

基于外观的方法通常提取和分析整个或特定面部区域的空间信息或时空信息。

*Yan et al. [21] proposed a framework of low-resolution FER based on image filter-based subspace learning (IFSL), including deriving discriminative image filters (DIFs), their combination, and an expression-aware transformation matrix.*

### *c)* 基于特征融合的 *FER (Feature Fusion-Based FER)*

这个融合是基于几何学的 *FER* 和外观的 *FER* 来融合来增强整个识别系统的鲁棒性，*Yao* 等人就利用 *MKL* 这一融合策略结合 2*D* 的纹理特征，3*D* 的形状特征和不同面部区

域的转换图谱做融合。

*Yao et al. [22] used the MKL fusion strategy to combine 2D texture features, 3D shape features, and their corresponding Fourier transform maps of different face regions.*

### d) 基于特征选择的 *FER (Feature Selection)*

特征选择是从给定的特征集合中选择一个相关且有用的子集，同时识别并去除冗余属性。

*Different from the spatial division with fixed grid, a hierarchical spatial division scheme (HSDS) [23] was proposed to generate multiple types of gradually denser grids and designed kernelized group sparse learning (KGSL) to learn a set of importance weights.*

关于上述四种不同 *FER* 方法中提到的代表性文献，也整理了对应的文献插图置于下图 *Figure 4* 中，供参考大致流程：

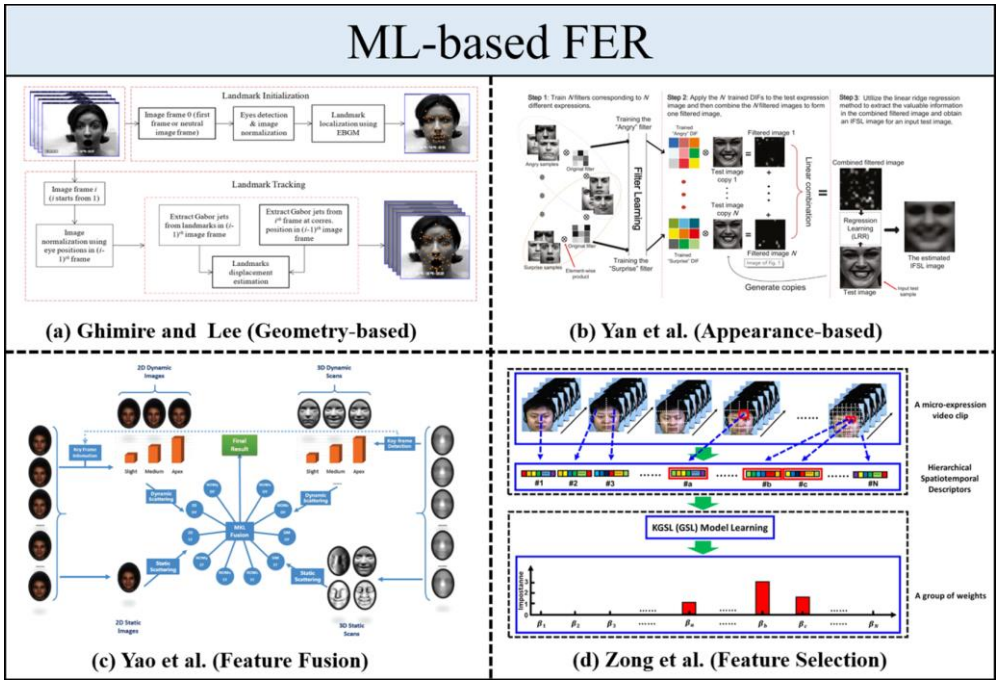



***Figure 4: Sample of ML-Based FER***

## 2. 基于深度学习的 *FER*（*DL-Based FER*）

现存的所有深度学习方法提取面部情感的骨干网络（*backbone networks*）都来自于知名的预训练卷积网络，例如：*VGG*、*VGG-face*，*ResNet*、*GoogLeNet*，考虑到不同的网络结构，还是把这一部分划分为了和之前一样的内容结构，即卷积网络、卷积-循环网络、对抗学习。

### *a)* 使用卷积网络学习的 *FER (ConvNet-Based FER)*

当使用相对较小的面部表情数据库时，基于 *convnet* 的 *FER* 经常设计转换学习或损失函数来克服过拟合。为了突出面部图像中最有用的信息，提出了各种注意机制来区分不同的特征。*Sun*[24]等人提出了一套新的知识迁移技术，包含一个深层的预训练教师模型和一个浅层的学生模型。

*Su et al. [24] proposed a novel knowledge transfer technique, which comprised a pre-trained deep teacher neural network and a shallow student neural network. Specifically, the AU-based model is trained on the residual network, which is then distilled and transferred for FMER.*

### b) 使用卷积-循环网络学习的 *FER (Conv-RNN-Based FER)*

对于动态的面部序列或视频，连续帧的时间相关性应被视为重要的线索。*RNN* 及其变体(*LSTM*)可以鲁棒地导出空间特征表示的时间特征。相比之下，代表性表达状态帧的空间特征可以通过 *CNN* 学习。基于 *ConvNet-RNN* 网络的架构，许多研究提出了级联融合或集成策略来捕获 *FER* 的空间和时间信息。

*The standard pipeline of ConvNet-RNN based FER using the ensemble strategy is to fuse outputs of two streams. For example, Zhang et al. [25] proposed a deep evolutional spatial-temporal network, which consists of a part-based hierarchical bidirectional recurrent neural network (PHRNN) and a multi-signal convolutional neural network (MSCNN), for analyzing temporal facial expression information and still appearance information, respectively.*

### *c)* 使用对抗学习的 *FER (Adversarial-Based FER)*

由于 GAN 可以生成不同姿态和视角下的合成面部表情图像，因此基于 GAN 的模型被用于姿态/视角不变的 FER 或身份不变的 FER。另外，现在也有了许多基于 GAN 的图像生成应用，不仅用于面部识别。

*For pose/view-invariant FER, Zhang et al. [26] proposed the GAN using AE structure to generate more facial images with different expressions under arbitrary poses and [27] further took the shape geometry into consideration.*

使用深度学习方法的面部表情识别的代表性文献同样整理后置于 Figure 8 中，供参考网络结构和流程。

由于视觉情感识别在识别人类情感方面的突出优势，大多数视觉情感识别研究都集

中在视觉情感识别上。然而，在某些环境下，当专用传感器无法捕获面部图像或仅捕获低分辨率的面部图像时，*FER* 将不适用。*EBGR* 旨在从全身视觉信息(如身体姿势)和身体骨架运动或上半身视觉信息(如手势、头部定位和眼球运动)中揭示一个人隐藏的情绪状态。*EBGR* 的一般流程包括人体检测(视为预处理)、特征表示和情感识别。从特征提取和情感识别过程是否端到端进行的角度出发，*EBGR* 系统分为基于 *ML* 的 *EBGR* 和基于 *DL* 的 *EBGR*。

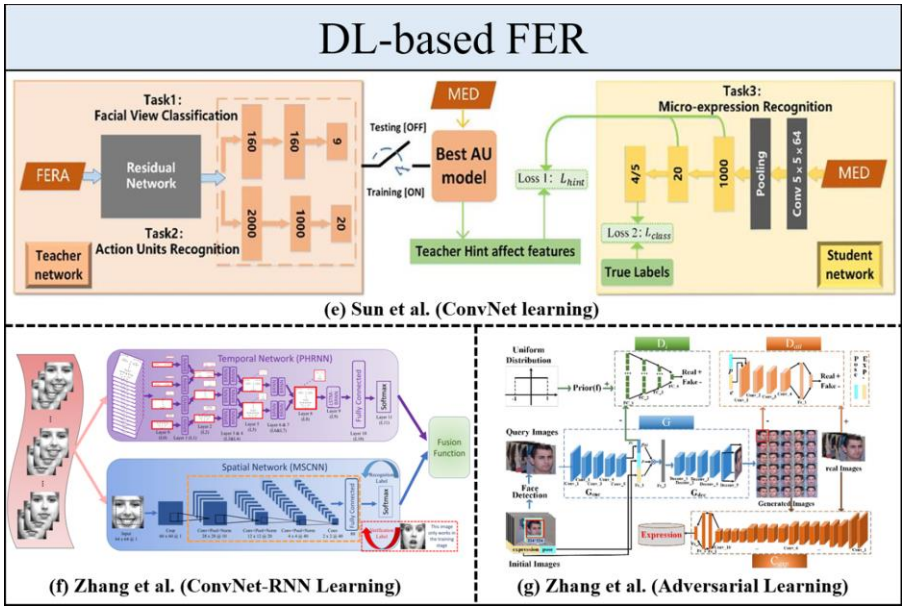



***Figure 5*: *Sample of DL-Based FER***

## 3. 基于机器学习的 *EBGR*（*ML-Based EBGR*）

***a)*** **基于统计方法或基于运动分析的 *EBGR (Statistic-Based or Movement-Based BGER)***

***b)*** **使用特征融合的 *EBGR (Feature Fusion)***

***c)*** **基于分类的 *EBGR (Classifier-Based EBGR)***

***d)*** **非运动状态下的 *EBGR (None acted EBGR)***

## 4. 基于深度学习的 *EBGR*（*DL-Based EBGR*）

尽管基于 *DL* 的 *EBGR* 系统不需要设计定制的特征提取器，但它们通常基于常用的姿态估计模型或低级特征提取器对输入数据进行预处理。通过基于 *CNN* 的网络、基于 *LSTM* 的网络或基于 *CNN-LSTM* 的网络，可以在空间、时间或时空维度上学习高级特征。许多研究已经证明了将不同的基于 *DL* 的模型与注意机制有效结合在一起的优势，

可以提高 *EBGR* 的性能。

这一部分的肢体情感识别分为两种方法：基于卷积-循环网络学习（*ConvNet-RNN learning*）的肢体情感识别；基于零样本学习（*Zero-Shot*）的肢体情感识别。

## 四、 生理情感分析（*Physiological-based emotion recognition*）

一个人的面部表情、文字、声音和肢体动作都可以很容易地收集到。由于身体信息的可靠性在很大程度上取决于社会环境和文化背景，以及测试者的性格，他们的情绪很容易被伪造。而生理信号的变化直接反映了人类情绪的变化，可以帮助人类识别、解读和模拟情绪状态。因此，通过生理信号来学习人类情绪是非常客观的。

基于生理的情感预测通常包括五个步骤：通过外界条件的变化来刺激被试的情绪→其次记录其生理上的信号→再通过生理信号预处理、特征分析、特征选择和还原提取特征→再再次训练分类模型，如 *SVM*、*KNN* 之类→最后基于离散情绪模型或维度情绪模型进行情感识别，这一过程如 *Figure* 6 所示。

*1) Stimulating subjects' emotions with images, music and videos; 2) Recording physiological signals that mainly include EEG, skin conductance, RESP, heart rate, EMG, and ECG; 3) Extracting features through physiological signals pre-processing, feature analysis, feature selection and reduction; 4) Training the classification model such as SVM, KNN, LDA, RF, NB and NN, etc.; and 5) Emotion recognition based on a discrete emotion model or a dimensional emotion model.*

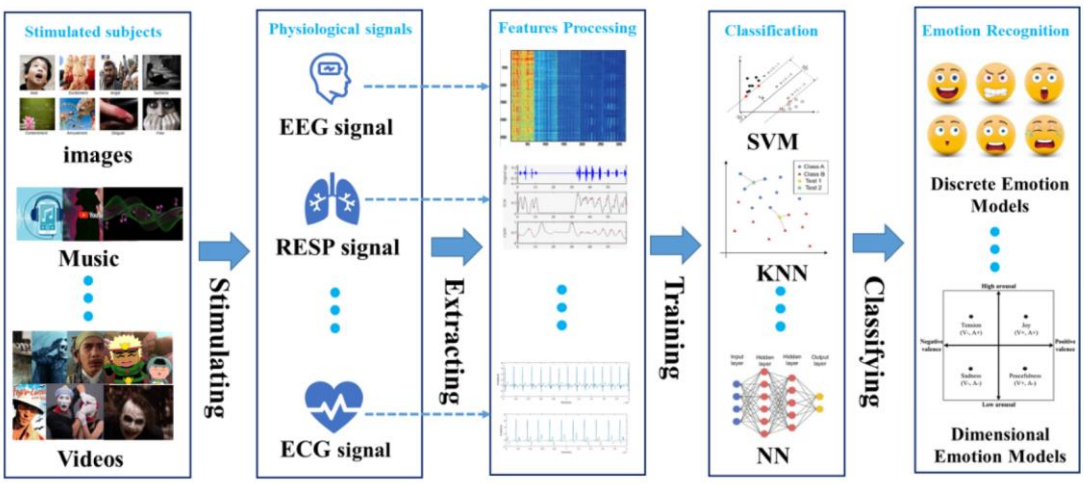



***Figure 6: Steps of Physiological-Based Emotion Recognition***

上述提及到的生理信号中，*EEG*（脑电图，*Electroencephalogram*）和 *ECG*（心电图，*Electrocardiogram*）可以提供简单、客观以及可信赖的数据用于识别被试情感，同时它们也是用于情感分析预测的最常用数据来源。跟上面一样，心脑电图也可以基于 *DL* 或

者 *ML* 来进行情感识别。

## (四) *Multi-Modal Affective Computing/Analysis*

根据融合策略分为特征级融合、决策级融合、模型级融合、混合融合；根据组合方式可以分为多物理模态情感分析、多生理模态情感分析、生理-物理模态情感分析，为避免由于翻译引起的各种歧义，这一部分的原文也附于下，以后看的时候先看英文！

*Nowadays, most reviews of multimodal affective analysis focus on multimodal fusion strategies and classify them into* ***feature-level fusion*** *(or early fusion),* ***decision-level fusion*** *(or late fusion),* ***model-level fusion****, and* ***hybrid-level fusion****. However, the multimodal affective analysis can be also varied with combinations of different modalities.*

*However, the multimodal affective analysis can be also varied with combinations of different modalities. Therefore, we categorize multimodal affective analysis into* ***multi-physical modality fusion*** *for affective analysis,* ***multi-physiological modality fusion*** *for affective analysis, and* ***physical-physiological modality fusion*** *for affective analysis, and further classify them based on four kinds of fusion strategies.*

关于四种不同融合策略的说明：

***Feature-level fusion***：*Feature-level fusion* 的特征融合提取来自多种模态的特征并最后生成一种通用的特征向量，随后将其送入分类器。

***Decision-level fusion***：*Decision-level fusion* 将来自所有模态的决策向量相互拼接，生成一种连接后的向量。

我个人补充：这两种向量分别又被称为 *early fusion* 和 *late fusion*，私以为这二者的区别就在于融合过程处于每个模态的特征提取和表示阶段的位置。之前看到的 *paper* 中提到，这两种方式都会引起不同模态间的交互冲突，由此又引出了一种新的融合策略，*intermediate fusion*？（好像是叫这个名字），不失为作为一种模型改进手段，将其与 *baseline* 比较。

***Model-level fusion***：这一融合方式发现了提取自不同模态间的特征的相关性，并使用、设计了具有松弛和平滑类型的融合模型，例如 *HMM* 或两阶段 *ELM*。

***Hybrid fusion***：结合了前两种的混合方式，下图 *Figure 7* 为采用这一融合策略的 *sentiment analysis* 模型结构。

多模态数据情感计算的每一部分看起来很困难，但实际上其实一点也不简单，所以为了保证我能顺利整理完，所以下面将只挑视觉-语音模态的情感计算进行整理，这也是

未来可能会使用最多的部分，剩余内容包括文本-语音和基于生理数据的情感计算如果后续需要参考，再回来看吧。

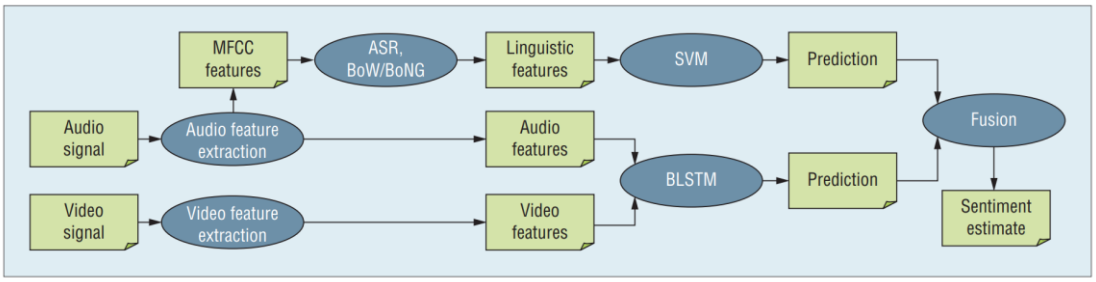



**Figure 1. System architecture for fusion of audio-visual and linguistic information (for in- and cross-domain analysis). Turnwise audio and video features are merged via early fusion and serve as input for the bidirectional long short-term memory (BLSTM) network, which in turn produces a sentiment prediction.**

***Figure 7*: *Sentiment Analysis for YouTube Videos***

使用注意力机制的视觉-语音情感计算。*Zhang*[29]等人使用了 *embedded attention*，以从各自模态中提取与情绪相关的区域，为更好融合两种不同模态特征，同时考虑到视频不同帧之间的表达差异，又提出了 *factorized bilinear pooling* (*FBP*)这一融合策略；*Zhao*[30]等人提出了一种新型的带有注意力模块和一致极性交叉熵损失函数（*polarity-consistent cross-entropy loss*）的视频音频注意力网络。*[Specifically, spatial, channel-wise, and temporal attentions are integrated with a 3D CNN for video frame segments, and spatial attention is integrated with a 2D CNN (ResNet-18) for audio MFCC segments.]*

使用卷积网络的视觉-语音情感计算。用 *CNN* 提取两种模态各自特征，然后用 *DBN* 做特征融合[31]；同时用 *2D* 和 *3D* 的 *CNN* 处理高层次预处理后的视频-音频信号表征，再用两阶段的 *ELM* 模型做特征融合，再用 *SVM* 做情感分类（计算）[32]。

## (五) *Related Datasets*

这一部分本来不想整理的，觉得如果要用就到 *Google* 那个专门的数据库搜集网站上搜一搜就有了，但想一想还是别懒了，就把名字 *ctrl cv* 过来，省的以后用的时候还麻烦。这里的分类和介绍可能不会做了，顶多分一分是带声的，还是带画的，如果真走投无路到这里找数据集了，估计我也是做不出来了罢。

- ***IEMOCAP***

*IEMOCAP*（*The Interactive Emotional Dyadic Motion Capture*）是一个情感分类对话数据集，包含了大约 12 个小时的多模态情感分类数据，通过对 10 名男演员和女演员基于剧本的即兴演出进行录制得到，每个场景包含两个说话人。

- ***MELD***

*MELD*（*Multimodal EmotionLines Dataset*）是一个情感分类对话数据集，来自对老友记剧集的标注，一个场景中最多包含 9 个说话人。相比于 *IEMOCAP*，该数据集规模更大。

● ***CMU-MOSEI/CMU-MOSI***

*CMU-MOSEI*（*CMU Multimodal Opinion Sentiment and Emotion Intensity* ）数据集是当前最大的情绪分析+情感分类（对话）数据集，所有场景只有一个说话人。由 *MultiComp Lab* 提供。

该数据集中包含了来自各个视频网站的单人说话视频，涵盖的主题包括新闻报道、英语教学、影片介绍、学校介绍、广告等。

此外该实验室还提供了一个相对较小的同类型数据集 *CMU-MOSI*（*The Multimodal Corpus of Sentiment Intensity*），不过只提供了 [-3,+3] 的情绪标注。

上面介绍的三个数据集都是三模态都有，此外还有一些双模态数据集，如 ***VideoEmotion*-8(*V+A*)**，***Ekman*-6(*V+A*)**，***RAVDESS*(*V+A*)**，***eNTERFACE*'05(*V+A*)** 等，具体介绍略。

就先整理这么多吧，实在感觉是用不到，剩下的到后面附的 *paper* 里去翻吧。

# (六) *Other Emotion Modelling Methods*

前面两部分其实提到了两种不同的情感分类建模方法，这一部分本来也是想偷懒放弃掉的，整理出来可能也太小众，不如前面两位重量级，但又一想，嘿，还是整理一下。

## 一、 维度情感模型（的补充）

同样是维度情感模型，再整理一下一维情感模型和三维情感模型。所谓一维情感模型不过就是正负两极。由于情感的快乐维度是个体情感的共有属性，许多不同的情感会借此相互制约，这还可以为个体情感的自我调节提供依据，但多数心理学家认为情感是由多个因素决定的，也因此产生后来的多维情感空间。

在三维情感模型中，除了考虑情感的极性和强度外，还有其他因素考虑到情感描述中。*PAD* 三维情感模型是当前认可度比较高的一种三维情感模型，该模型定义情感具有愉悦度、唤醒度、和优势度三个维度，其中 *P* 代表愉悦度，表示个体情感状态的正负特性；*A* 代表唤醒度，表示个体的神经生理激活水平；*D* 代表优势度，表示个体对情景和他人的控制状态。

另外，还有 *APA* 三维情感空间模型，该模型采用亲和力、愉悦度和活力度三种情感

属性，能够描述绝大多是情感。

除了以上三种情感模型外，还有更复杂的情感模型。心理学家 *Izard* 的思维理论认为情绪有愉悦度、紧张度、激动度和确实度 4 个维度。愉悦度代表情感体验的主观享乐程度，紧张度和激动度代表人体神经活动的生理水平，确信度代表个体感受情感的程度。

心理学家 *Krech* 认为情感的强度是指情感具有由弱到强的变化范围，同时还以紧张水平、复杂度、快乐度 3 个指标来进行量化。紧张水平是指对要发生的事情的事先冲动，复杂度是对复杂情感的量化，快乐度是表示情感所处的愉快和不愉快的程度，故可以从这四个维度来判断人的情感。

另外，心理学家 *Frijda* 提出了情感具有愉快、激活、兴趣、社会评价、惊奇和复杂共 6 个维度的观点，但高维情感空间的应用存在较大难度，因此在实际中很少使用。

维度情感模型是用人类情感体验的欧氏距离空间描述，其主要思想是人类的所有情感都涵盖于情感模型中，且情感模型不同维度上的不同取值组合可以表示一种特定的情感状态。虽然维度情感模型是连续体，基本情感可以通过一定方法映射到情感模型上，但对于基本情感并没有严格的边界，即基本情感之间可以逐渐、平稳转化。维度情感模型的发展为人类的情感识别、情感合成和调节提供了模型基础。

## 二、 其他情感模型

### 1. *OCC* 情感模型

该模型是针对情感研究而提出的最完整的情感模型之一，它将 22 种基本情感根据其起因分为三类：事件的结果、仿生代理的动作和对于对象的观感，并对这三类定义了情感的层次关系，可以描述特定情感的产生条件和后续发展。*OCC* 模型给出了各类情感产生的认知评价方式。同时，该模型根据假设的正负极性和个人对刺激事件反应是否高兴、满意和喜欢的评价倾向构成情感反应。

### 2. *HMM* 情感模型

隐马尔可夫模型情感模型，该模型有三种情感状态，分别是感兴趣、高兴、悲伤，并且可根据需要扩展到多种情感状态。在模型中，情感状态是通过观测到如情绪响应上升时间、峰值间隔的频率变化范围等情感特征得到的，并通过转移概率来描述情感状态之间的相互转移，从而输出一种最可能的情感状态。

该模型适合表现由不同情感组成的混合情感，如忧伤可以由爱和悲伤组成。另外，

还适合表现由若干单一的情感状态基于时间的不断交替出现而成的混合情感，如爱恨交织的情感状态就可能是爱恨两种之间循环。该模型的不足之处在于，对于相同的刺激，其感知结果是确定的。

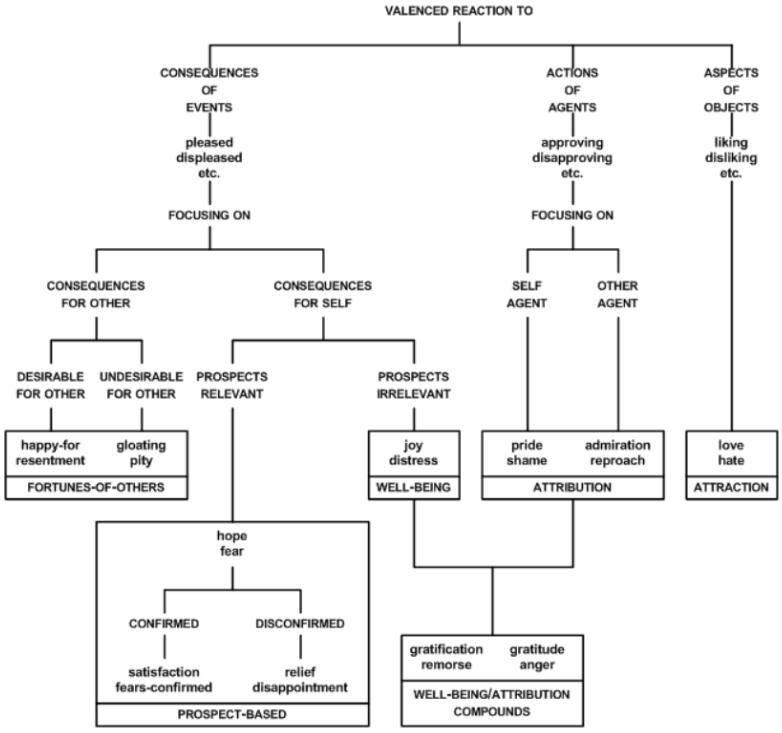



***Figure 8*: *OCC Emotion Model***

### 3. 分布式情感模型

该模型是针对外界刺激建立起来的一种分布式情感模型，整个分布式系统是将特定的外界情感事件转换成与之相对应的情感状态，过程分为以下两个阶段：

1、由事件评估器评价事件的情感意义，针对每一类相关事件，分别定义一个事件评估器，当事件发生时，先确定事件的类型和信息，然后选择相关事件评估器进行情感评估，并产生量化结果情感脉冲向量 *EIV*。

2、对 *EIV* 归一化得到 *NEIV*，通过情感状态估计器 *ESC* 计算出新的情感状态。事件评估器、*EIV*、*NEIV* 及 *ESC* 均采用神经网络实现。

## (七) *Extension about Discrete Emotion*

这一部分还是懒得整理来着，但可能未来做离散情感分类需要依照这个找单独的情感数据集，所以还是按照现存的一些分类体系来整理一下，会包括很多种不同情感的分类或者情感之间关系。

## 一、 基本情绪（*Basic Emotion*）

在 1890 年，威廉·詹姆斯根据身体参与程度提出了四种基本情绪：恐惧（*fear*）、悲痛（*grief*）、爱（*love*）和盛怒（*rage*）。

*Paul Ekman* 确定了六种基本情绪：愤怒（*anger*）、厌恶（*disgust*）、恐惧（*fear*）、幸福（*happiness*）、悲伤（*sadness*）和惊喜（*surprise*）。情绪可以与面部表情有关。在 20 世纪 90 年代，*Ekman* 提出了一个扩大版的基本情绪列表，包括一系列并非都编码在面部肌肉中的积极和消极情绪。新包含的情绪是：娱乐（*amusement*）、蔑视（*contempt*）、自满（*contentment*）、尴尬（*embarrassment*）、兴奋（*excitement*）、内疚（*guilt*）、成就中的自豪（*pride in achievement*）、轻松（*relief*）、满足（*satisfaction*）、感官愉悦（*sensory pleasure*）和羞耻（*shame*）。

*Richard* 和 *Bernice Lazarus* 在 1996 年将情绪列表扩大到 15 种情绪，在《*Passion and Reason*》一书中列出：审美体验（*aesthetic experience*）、愤怒（*anger*）、焦虑（*anxiety*）、同情（*compassion*）、抑郁（*depression*）、嫉妒（*envy*）、惊吓（*fright*）、感激（*gratitude*）、内疚（*guilt*）、幸福（*happiness*）、希望（*hope*）、嫉妒（*jealousy*）、爱（*love*）、自豪（*pride*）、轻松（*relief*）、悲伤（*sadness*）和羞耻（*shame*）。

加州大学伯克利分校的研究人员确定了 27 类情感：钦佩（*admiration*）、崇拜（*adoration*）、审美欣赏（*aesthetic appreciation*）、娱乐（*amusement*）、愤怒（*anger*）、焦虑（*anxiety*）、敬畏（*awe*）、尴尬（*awkwardness*）、厌倦（*boredom*）、冷静（*calmness*）、困惑（*confusion*）、渴望（*craving*）、厌恶（*disgust*）、共情之痛（*empathic pain*）、魅惑（*entrancement*）、兴奋（*excitement*）、恐惧（*fear*）、恐怖（*horror*）、兴趣（*interest*）、快乐（*joy*）、怀旧（*nostalgia*）、轻松（*relief*）、浪漫（*romance*）、悲伤（*sadness*）、满足（*satisfaction*）、性欲（*sexual*）和惊喜（*surprise*），这一体系的提出是基于 2185 个旨在激发某种情绪的短视频。

## 二、 相对基本情绪（*Contrasting basic emotions*）

*A 2009 review[33] of theories of emotion identifies and contrasts fundamental emotions according to three key criteria for mental experiences that:*

1. *have a strongly motivating subjective quality like pleasure or pain;*
2. *are a response to some event or object that is either real or imagined;*
3. *motivate particular kinds of behavior.*

*The combination of these attributes distinguishes emotions from sensations, feelings and moods.*

## 三、 基于 *HUMAINE* 的情感分类方法

*HUMAINE*（*Human-Machine Interaction Network on Emotion*）提出的 *EARL*（*emotion annotation and representation language*）对 48 中情绪进行了分类：

- 消极且有力的（*Negative and forceful*）：愤怒（*Anger*）、烦恼（*Annoyance*）、蔑视（*Contempt*）、厌恶（*Disgust*）、激怒（*Irritation*）
- 消极且不在掌控中的（*Negative and not in control*）：焦虑（*Anxiety*）、尴尬（*Embarrassment*）、恐惧（*Fear*）、无助（*Helplessness*）、无力（*Powerlessness*）、担忧（*Worry*）
- 消极思维（*Negative thoughts*）：自豪（*Pride*）、怀疑（*Doubt*）、嫉妒（*Envy*）、挫折（*Frustration*）、内疚（*Guilt*）、羞耻（*Shame*）
- 消极且被动的（*Negative and passive*）：厌倦（*Boredom*）、绝望（*Despair*）、失望（*Disappointment*）、伤害（*Hurt*）、悲伤（*Sadness*）
- 忧虑（*Agitation*）：压力（*Stress*）、震惊（*Shock*）、紧张（*Tension*）
- 积极且有活力的（*Positive and lively*）：娱乐（*Amusement*）、高兴（*Delight*）、得意（*Elation*）、兴奋（*Excitement*）、幸福（*Happiness*）、快乐（*Joy*）、愉快（*Pleasure*）
- 关心（*Caring*）：爱慕（*Affection*）、共情（*Empathy*）、友谊（*Friendliness*）、爱（*Love*）
- 积极思维（*Positive thoughts*）：勇气（*Courage*）、希望（*Hope*）、谦逊（*Humility*）、满足（*Satisfaction*）、信任（*Trust*）
- 平静的积极（*Quiet positive*）：冷静（*Calmness*）、自满（*Contentment*）、放松（*Relaxation*）、轻松（*Relief*）、平静（*Serenity*）
- 反应（*Reactive*）：兴趣（*Internet*）、礼貌（*Politeness*）、惊喜（*Surprise*）

## 四、 按组分类的 *Parrott* 情绪分类

*Shaver*[34]等人提出了树状结构的情绪列表，并随后在 *Parrott*[35]中再次出现。

## 五、 *Pultchik* 的情绪轮子（*Plutchik's wheel of emotions*）

*In 1980, Robert Plutchik diagrammed a wheel of eight emotions: joy, trust, fear, surprise, sadness, disgust, anger and anticipation, inspired by his Ten Postulates.[36][37] Plutchik also theorized twenty-four "Primary", "Secondary", and "Tertiary" dyads (feelings composed of two*

*emotions).*[38][39][40][41][42][43][44] *The wheel emotions can be paired in four groups:*

其实就是将最开始提到的八种情绪做三级细分，可以用下图表示：

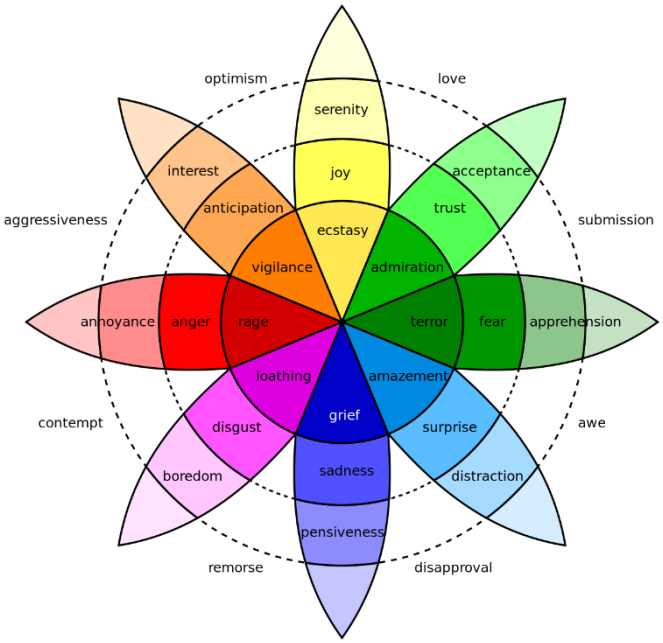



*Figure 9: Plutchik-Wheel*

每一级由内到外分别对应不同程度，最内程度最深，向外就逐层递减，然后不同相邻轮子之间可以产生类似于情绪加和关系，对角向的轮子之间是对立关系，比如：

*Primary dyad = one petal apart = Love = Joy + Trust*

*Secondary dyad = two petals apart = Envy = Sadness + Anger*

*Tertiary dyad = three petals apart = Shame = Fear + Disgust*

*Opposite emotions = four petals apart = Anticipation ∉ Surprise*

可以这么表示：

经过这一运算后就会产生更加复杂的配对组合情绪，会有 24 个二元配对和 32 个三元配对组合情绪，使 56 种情绪处于一个强度水平（*intensity level*）。组合后的情感会根据其组合前的基础情感等产生更加复杂的情感关系，如相反情绪、基础相反情绪等，也建议自行查阅文献。

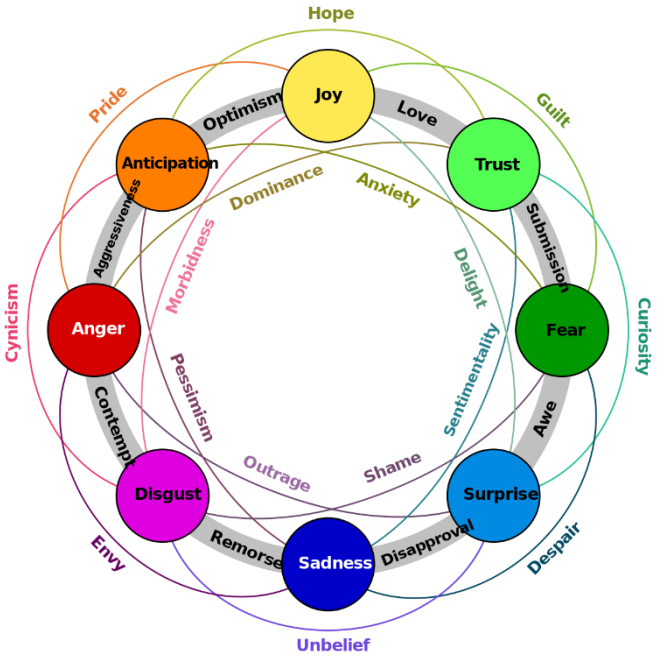



*Figure 10*: *Plutchik Dyads*

## 六、 六维情绪轴（*Six emotion axes*）

好像标题翻译错了，反正原文的说法也粘贴到后面了，用的时候再看吧。

*MIT researchers [45] published a paper titled "An Affective Model of Interplay Between Emotions and Learning: Reengineering Educational Pedagogy—Building a Learning Companion" that lists six axes of emotions with different opposite emotions, and different emotions coming from ranges.*

# (八) *Extension about multimodal*

这一部分可能主要还是关于模态融合的东西，关于 *early fusion* 和 *late fusion*，了不起再加上我看到的一篇 *affect computing* 和自动驾驶的内容，再加上我用到的图片整体情感的多模态解决方法，要真的都加上了就很了不起了！

## 一、 模态间的不同融合策略

多模态 *fusion* 是一种处理多模态数据的技术，它可以将来自不同传感器或模态的信息进行整合和融合。在现实生活中，我们接收到的信息往往是多模态的，例如图像、文本、语音等。这些不同模态的数据包含了丰富的信息，但各自的表达方式和特点也不同，因此需要将它们有效地结合起来，以提供更全面、准确和可靠的分析和决策支持。多模

态 *fusion* 的目标是将多个模态的数据融合成一个统一的表示或特征向量，使得整个系统能够对多模态数据进行综合分析、理解和推理。通过融合不同模态的信息，可以弥补单一模态的局限性，并提高对复杂场景和任务的处理能力。

根据前面提到的不同 *fusion* 方法，按照不同时机分为了 *early*、*late* 和 *intermediate*，所以下面也就按照这几种方法进行分别介绍。

### 1. *Early fusion*（*feature-level fusion*）

*Early fusion* 将多个独立的数据集融合成一个单一的特征向量，然后输入到机器学习分类器中。由于多模态数据的 *early fusion* 往往无法充分利用多个模态数据间的互补性，且 *early fusion* 的原始数据通常包含大量的冗余信息。因此，*early fusion* 方法常常与特征提取方法相结合以剔除冗余信息，如主成分分析（*PCA*）、最大相关最小冗余算法（*mRMR*）、自动解码器（*Autoencoders*）等。

### 2. *Late fusion*（*decision-level fusion*）

*late fusion* 则是将不同模态数据分别训练好的分类器输出打分(决策)进行融合。这样做的好处是，融合模型的错误来自不同的分类器，而来自不同分类器的错误往往互不相关、互不影响，不会造成错误的进一步累加。常见的 *late fusion* 方式包括最大值融合(*max-fusion*)、平均值融合(*averaged-fusion*)、 贝叶斯规则融合(*Bayes' rule based*)以及集成学习(*ensemble learning*)等。其中集成学习作为 *late fusion* 方式的典型代表，被广泛应用于通信、计算机识别、语音识别等研究领域。

### 3. *Intermediate fusion*

*intermediate fusion* 是指将不同的模态数据先转化为高维特征表达，再于模型的中间层进行融合。以神经网络为例，*intermediate fusion* 首先利用神经网络将原始数据转化成高维 特征表达，然后获取不同模态数据在高维空间上的共性。*intermediate fusion* 方法的一大优势是可以灵活的选择融合的位置。关于这一融合策略也有很多不同融合方法：

基于简单操作的融合：来自不同的模态的特征向量可以通过简单地操作来实现整合，比如拼接（*Concatenation*）和加权求和（*Weighted Sum*）。这样的简单操作使得参数之间的联系几乎没有，但是后续的网络层会自动对这种操作进行自适应。

基于注意力机制的融合办法：很多的注意力机制已经被应用于融合操作了。注意力机制通常指的是一组“注意”模型在每个时间步动态生成的一组标量权重向量的加权和。

这组注意力的多个输出头可以动态产生求和时候要用到的权重，因此最终在拼接时候可以保存额外的权重信息。在将注意机制应用于图像时，对不同区域的图像特征向量进行不同的加权，得到一个最终整体的图像向量。

基于双线性池化的融合办法（基于张量的融合方法）：双线性池化主要用于融合视觉特征向量和文本特征向量来获得一个联合表征空间，方法是计算他们俩的外积，这种办法可以利用这俩向量元素的所有的交互作用，也被称作 *second-order pooling*。

## 二、 基于 *ANP* 的 *text-vision* 模态情感计算

由于这是我用过的一种方法，所以我可能在这里不会写的特别详细，但应该会在这个字[46]后面的右上方加一个参考文献的出处，如果这篇文章被传阅了，并且有地方看不懂，可以来问我，但整理应该是做不到了。

首先贴出来这个神经网络的结构图吧，总体来讲其实还是一个卷积神经网络，不过一眼就能看出来的，和最基础的 *CNN* 不一样地方在于它没有加最后的全连接层，而是选用了一种其他的拼接方法做替代，想必也是做了实验才改进的。

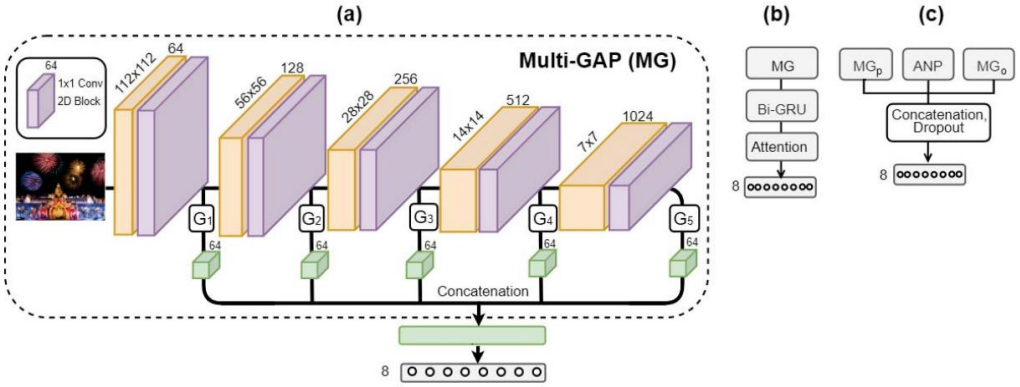



**Fig. 2**: (a) The proposed Multi-GAP (MG) base network which extracts features from each network block. The GAP layers (G) reduce the overall network parameters, making it less susceptible to over-fitting. Two flavours are explored: (b) MG with feature dependency (Bi-GRU + Attention). (c) Fusion of MG features ($MG_p$ and $MG_o$) and other auxilliary features (ANP).

***Figure 11*: *MultiGAP Sentiment Analysis***

这个结构很简单就能看出来，前面的黄色 *conv* 应该是用来做特征提取，后面的紫色 *conv* 是用来做通道降维，没有用 *FC*，拿 *GAP* 来代替，最后将多层的 *GAP* 输出经过 *Concatenation* 得到最后的某一模态的表征向量。

*MG* 网络的输出被看作具有 *T* 个时序步骤（*T*=*N*）的序列，也即一种序列数据，在这种情况下，特征从初期到后期的层次之间的转变可以通过 *Bi-GRU* 网络学习。在双向的 *GRU* 后面又加了一个 *soft attention mask*，用于强化重要特征，弱化无关特征。

关于软注意力机制：软注意力机制用于在一个序列或集合中选择重要的部分。它主要是通过给不同的

部分分配不同的权重来实现这个目标。与硬注意力机制不同，软注意力机制可以给一个序列或集合中的每个元素分配一个处于 0-1 之间的权重，而不是单单仅对元素做 0-1 选择。

关于注意力机制其实也可以再讲一讲，一般来讲的注意力机制可以分为三种，除了上面提到的软注意力（*soft attention*）机制，当然就有硬注意力机制，还有一个自注意力机制，这哥仨的区别在于：

1. *Soft*/*Global Attention*(软注意机制)：对每个输入项的分配的权重为 0-1 之间，也就是某些部分关注的多一点，某些部分关注的少一点，因为对大部分信息都有考虑，但考虑程度不一样，所以相对来说计算量比较大。

2. *Hard*/*Local Attention*(硬注意机制)：对每个输入项分配的权重非 0 即 1，和软注意不同，硬注意机制只考虑那部分需要关注，哪部分不关注，也就是直接舍弃掉一些不相关项。优势在于可以减少一定的时间和计算成本，但有可能丢失掉一些本应该注意的信息。

3. *Self*/*Intra Attention*(自注意力机制)：对每个输入项分配的权重取决于输入项之间的相互作用，即通过输入项内部的"表决"来决定应该关注哪些输入项。和前两种相比，在处理很长的输入时，具有并行计算的优势。

*Fusion schemes are common strategies that can be exploited to accumulate information from multiple feature descriptors or classifiers to further create more robust frameworks. In our work, we explore both early (feature-level) and late (output-level) fusion. In contrast to the work of, we extend our MG network into multiple streams so that the fusion schemes can be exploited to take advantage of information learned through different networks. Specifically, for early fusion, we use features derived from the Object MG network and Places MG network, together with high-level Adjective-Noun Pair (ANP) features (shown in Figure 2(c)). This process merges the bottlenecked features (without the classifier head) of the networks (except the ANP which are merely features). Meanwhile, late fusion is performed by averaging the predicted softmax values of all classifiers; the class with the highest averaged value is chosen as the predicted class. Note that the ANP features are put through a MLP network with 2 hidden layers (1024 and 512 nodes) and a 8-node output layer.*

反正总之就是在图片中提取了图片中存在的物体、场景等信息，再结合 *ANP*，并使用 *late fusion* 的模态间融合策略完成了最终的情感分类，上面摘的英文的模态融合部分的解释，也没有他说的那么复杂，看我这个大概的解释也是那么个意思。

## 三、 情感计算/心理波动现阶段应用

这应该是最后一部分了！本来这部分是真的不用整理的，因为也真的没有什么用，但是看资料感觉这里其实还是蛮有意思，就真的只做一些 *ctrl c+v* 的工作，毕竟这里我

是真的不懂。

这一部分的应用首先就是在机器人，包括声音合成，面部表情。

语音是表达情感的主要方式之一，因为我们人类总是能够通过他人的语音轻易地判断他人的情感状态。语音的情感主要表现在两个部分，一个是语音中所包含的语言内容，另一个是声音本身所具有的特征，比如音调的高低变化等。我们可以利用特定的声音风格加上文字内容合成语音，便可以表达特定的情感，带有情感的语音可以让消费者在使用的时候感觉更人性化、更温暖。

目前的语音合成通常都是通过将需要合成的文字内容和特定风格的语音输入到神经网络中，然后让神经网络合成特定风格的语音。然而，目前的神经网络无法高效地将语音内容和风格分解。如下图所示，微软的研究者在最近提出利用博弈论中对抗和合作的思想来生成特定风格的语音数据，这个模型能够有效地将语音内容和风格分解，从而使得在语音生成方面风格可控，该模型在风格迁移、情感建模等任务上均取得了不错的进展。

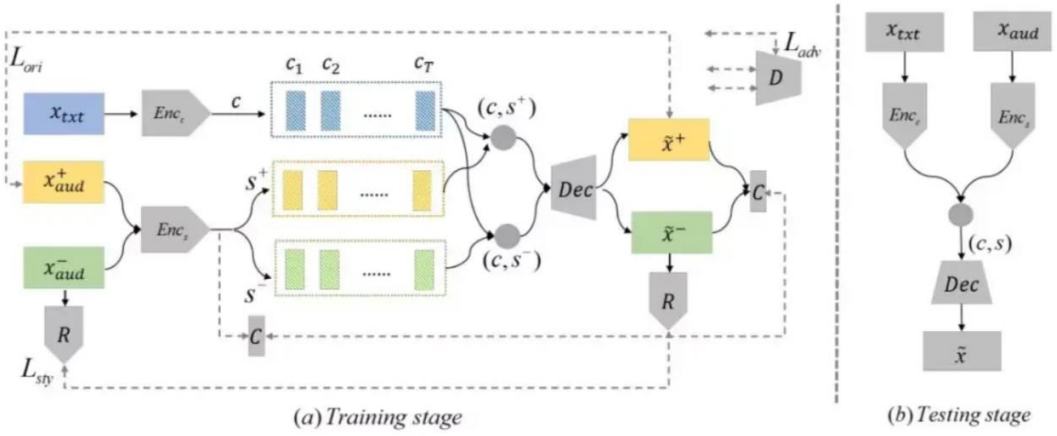



***Figure 12*: *Training Step for Video Generation***

面部表情是表现情感的一个重要途径，主要通过脸部、眼睛或者肌肉位置的变化来表达情感。不同国家的人面部表情各不相同，亚洲人民的面部表情的强度相对较低，因为在亚洲文化中，面部表现出一些特殊情绪是不礼貌的。

面部表情的生成是一项具有挑战性的任务，因为它需要对输入面部图像进行高级语义理解。在传统方法中，合成的面部分辨率通常很低。目前主要流行基于深度学习的方法进行面部表情图像生成，比如有研究利用生成对抗网络（*GAN*）进行带有指定情绪的面部表情生成，如图所示（摘自 *ExprGAN*），该模型可用于可控表情的面部表情生成，可以很好地表达不同的情感。

一个更为综合的情感表达的例子是对话系统，图灵在 1950 年就提出了著名的图灵

测试，他认为如果一台机器能够与人类展开对话而不能被辨别出其机器身份，那么称这台机器具有智能。我们在文章的开头谈到，如果机器不具有情感表达，那么人们可能会认为机器一点都不够智能。因此在与机器进行对话时，机器能够识别和表达情感是一件非常重要的事情。来自哈佛和微软的研究者们就尝试着让对话机器人能够综合语言信息和视觉信息进行带有情感表达的对话，针对问题「*Did you have a good time*?」，对话机器在看到不同的视觉场景会有不同情感表达。当图像是一个抬头并带有笑脸的小男孩，因此机器会回复「*We had a great time at the beach*!」，而当图像是一个低头的小女孩，机器会回复「*She just hates going for a walk*!」。

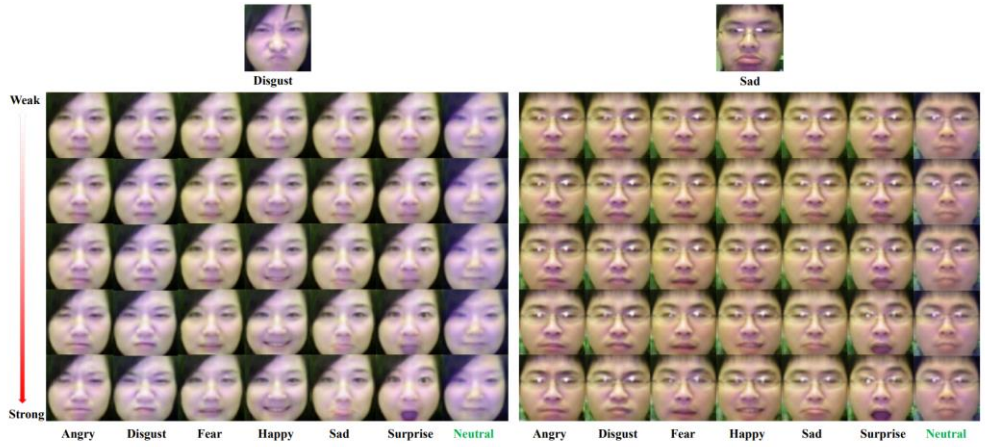



Figure 3: Face images are transformed to new expressions with different intensity levels. The top row contains the input faces with the original expressions, and the rest rows are the synthesized results. Each column corresponds to a new expression with five intensity levels from weak to strong. The *Neutral* expression which is not in the training data is also able to be generated.

***Figure 13*: *GAN for Face Generation***

另一个被举例的应用场景是驾驶，尽管这一资料上把它叫做决策。

大量的研究表明，人在解决某些问题的时候，纯理性的决策过程并不是一个最优解，在决策的过程中，如果有生理反应（如情感）加入到决策过程中，这有可能帮助我们找到更优的解。如果我们将情感机制纳入到强化学习算法的设计当中，那么智能体（*Agent*）会发什么有趣的事情？

举个例子，我们人类在遇到不利于我们生存的情况下，我们的交感神经系统（*Sympathetic Nervous System*, *SNS*）会分泌一系列激素促使我们的心跳、血压以及肾上腺素升高，并导致我们产生恐惧的情绪，这种恐惧的情绪会加速我们对风险规避的学习。如果我们将这种恐惧情绪加入到强化学习的智能体并辅助智能体决策，智能体在探索效率上可能会发生一定的变化。

微软的研究者在这个问题上给出了自己的答案，他们提出了一种基于周围血管搏动测量（*Peripheral Pulse Measurements*）的内在奖励的强化学习新方法，这种内在奖励是

与人类神经系统的响应相关的。作者的假设是这种奖励函数可以帮助强化学习解决稀疏性（*sparse*）和倾斜性（*skewed*），以此提高采样效率。

汽车驾驶是一个生活中很常见的任务，这既依赖于内部的奖励，也依赖于外表的奖励。当我们在高速驾驶汽车的时候，我们的神经系统是高度激活的，这有助于我们应对驾驶过程中出现的突发状况，比如需要紧急调整方向来防止撞到突然走向道路中间的行人以避免事故。因此，当遇到突发情况时，这种生理内部的反馈会有助于我们更好地评估当前的环境并帮助我们做出有利的决策。

如图所示，与一般强化学习模型的不同之处在于，作者提出的强化学习模型的奖励主要分为两个部分，一个是外部环境的奖励（*Extrinsic Reward*），一个是由内部生理反应产生的内部奖励（*Intrinsic Reward*）。作者利用皮肤周围血管血液体积，即比如血容量脉搏波动（*Blood Volume Pulse Wave*），来模拟内部生理状态的反应。核心思想是如果人在遇到某种紧急的情况，那么人的紧张情绪就会通过生理反应表现出来，比如血容量脉搏波动变大。

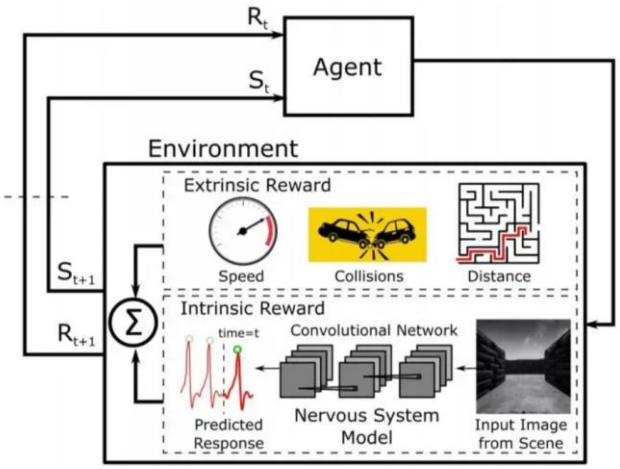
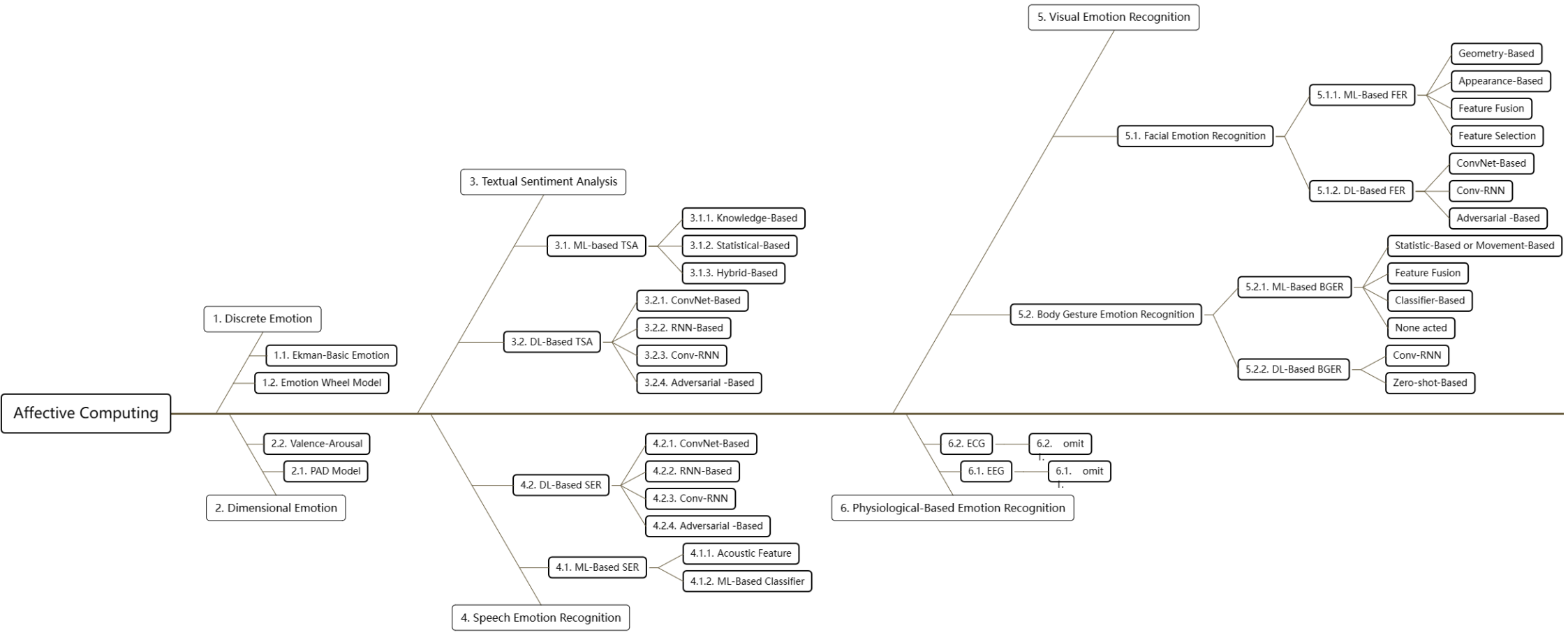
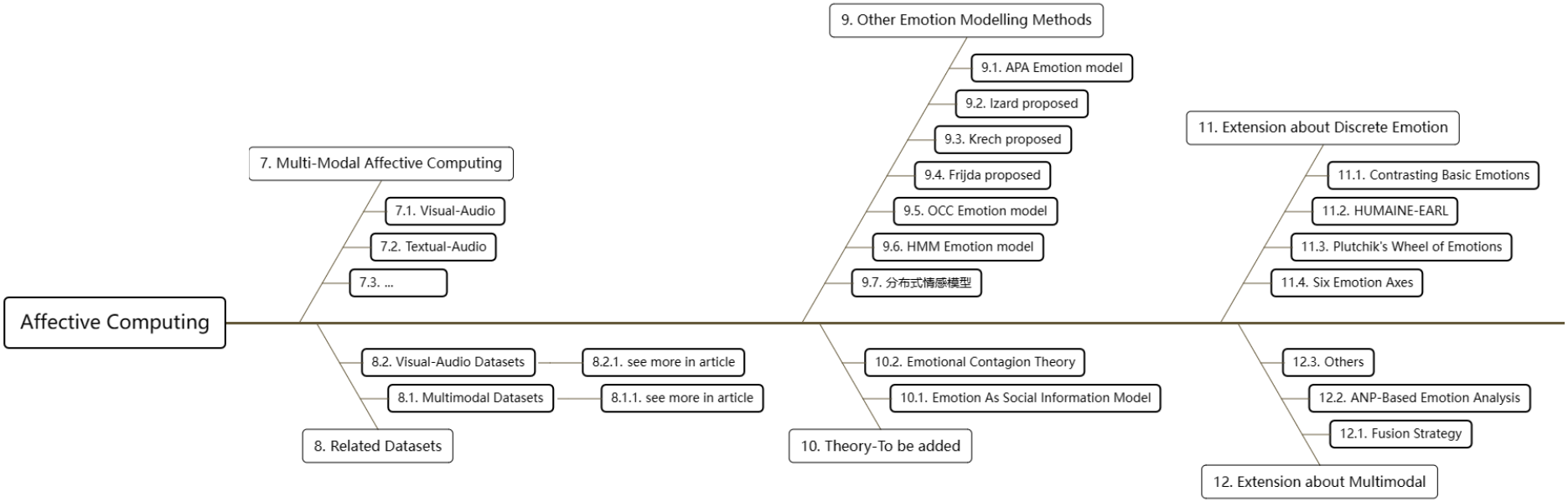



***Figure 14*: *The Figure I do not Understand***

这种方法的关键问题之一就是如何确定在开车过程中哪种驾驶场景会导致驾驶者心理出现波动，比如心跳加快和血压升高。作者找了四个人来获取皮肤周围血液体积的变化，具体做法是让这四个人分别在这个模拟驾驶场景中进行驾驶，并记录每一帧图像（驾驶场景）的变化以及参与者本人对应的血容量脉冲波动数据。作者利用获取到的数据对一个八层的卷积神经网络进行训练，图像帧作为输入数据，血容量脉冲波动作为标签，值在 0 到 1 之间。训练好的模型便可用来预测特定驾驶场景的心理反应，这种心理反应就是我们前面提到的内部奖励。


Affective Computing
1. Discrete Emotion
1.1. Ekman-Basic Emotion
1.2. Emotion Wheel Model
2. Dimensional Emotion
2.2. Valence-Arousal
2.1. PAD Model
3. Textual Sentiment Analysis
3.1. ML-based TSA
3.1.1. Knowledge-Based
3.1.2. Statistical-Based
3.1.3. Hybrid-Based
3.2. DL-Based TSA
3.2.1. ConvNet-Based
3.2.2. RNN-Based
3.2.3. Conv-RNN
3.2.4. Adversarial -Based
4. Speech Emotion Recognition
4.2. DL-Based SER
4.2.1. ConvNet-Based
4.2.2. RNN-Based
4.2.3. Conv-RNN
4.2.4. Adversarial -Based
4.1. ML-Based SER
4.1.1. Acoustic Feature
4.1.2. ML-Based Classifier
5. Visual Emotion Recognition
5.1. Facial Emotion Recognition
5.1.1. ML-Based FER
Geometry-Based
Appearance-Based
Feature Fusion
Feature Selection
5.1.2. DL-Based FER
ConvNet-Based
Conv-RNN
Adversarial -Based
5.2. Body Gesture Emotion Recognition
5.2.1. ML-Based BGER
Statistic-Based or Movement-Based
Feature Fusion
Classifier-Based
None acted
5.2.2. DL-Based BGER
Conv-RNN
Zero-shot-Based
6. Physiological-Based Emotion Recognition
6.2. ECG
6.2. omit 1.
6.1. EEG
6.1. omit 1.


Affective Computing
7. Multi-Modal Affective Computing
7.1. Visual-Audio
7.2. Textual-Audio
7.3. ...
8. Related Datasets
8.2. Visual-Audio Datasets
8.2.1. see more in article
8.1. Multimodal Datasets
8.1.1. see more in article
9. Other Emotion Modelling Methods
9.1. APA Emotion model
9.2. Izard proposed
9.3. Krech proposed
9.4. Frijda proposed
9.5. OCC Emotion model
9.6. HMM Emotion model
9.7. 分布式情感模型
10. Theory-To be added
10.2. Emotional Contagion Theory
10.1. Emotion As Social Information Model
11. Extension about Discrete Emotion
11.1. Contrasting Basic Emotions
11.2. HUMAINE-EARL
11.3. Plutchik's Wheel of Emotions
11.4. Six Emotion Axes
12. Extension about Multimodal
12.3. Others
12.2. ANP-Based Emotion Analysis
12.1. Fusion Strategy

## *Reference Paper*


[1] *J. Chen*, *H. Huang*, *S. Tian*, *Y. Qu*, "*Feature selection for text classification with Naïve Bayes*," *Expert Syst. Appl.*, *vol.* 36, 5432–5435, 2009, https://doi.org/10.1016/j.eswa.2008.06.054.

[2] *T. Le*, "*A hybrid method for text-based sentiment analysis*," *in Proc.* 2019 *Int. Conf. Comput. Sci. Comput. Intell. CSCI*, 2019, *pp.* 1392–1397, https://doi.org/10.1109/CSCI49370.2019.00260.

[3] *D. Li*, *R. Rzepka*, *M. Ptaszynski*, *K. Araki*, "*A novel machine learning-based sentiment analysis method for Chinese social media considering Chinese slang lexicon and emoticons*," *in Honolulu*, *Hawaii*, *USA*, 2019.

[4] *R. Yin*, *P. Li*, *B. Wang*, "*Sentiment lexical-augmented convolutional neural networks for sentiment analysis*," *in Proc.* 2017 *IEEE Second Int. Conf. Data Sci. Cyberspace DSC*, 2017, *pp.* 630–635, https://doi.org/10.1109/DSC.2017.82.

[5] *A. Conneau*, *H. Schwenk*, *L. Barrault*, *Y. Lecun*, "*Very deep convolutional networks for text classification*," *in Proc.* 15*th Conf. Eur. Chapter Assoc. Comput. Linguist. Vol.* 1 *Long Pap.*, *Association for Computational Linguistics*, *Valencia*, *Spain*, 2017, *pp.* 1107–1116. https://www.aclweb.org/anthology/E17-1104 (*accessed August* 13, 2020).

[6] *R. Johnson*, *T. Zhang*, "*Deep pyramid convolutional neural networks for text categorization*," *in Proc.* 55*th Annu. Meet. Assoc. Comput. Linguist. Vol.* 1 *Long Pap.*, *Association for Computational Linguistics*, *Vancouver*, *Canada*, 2017, *pp.* 562–570, https://doi.org/10.18653/v1/P17-1052.

[7] *B. Huang*, *K. Carley*, "*Parameterized convolutional neural networks for aspect level sentiment classification*," *in Proc.* 2018 *Conf. Empir. Methods Nat. Lang. Process.*, *Association for Computational Linguistics*, *Brussels*, *Belgium*, 2018, *pp.* 1091–1096, https://doi.org/10.18653/v1/D18-1136.

[8] *Z.-Y. Dou*, "*Capturing user and product information for document level sentiment analysis with deep memory network*," *in Proc.* 2017 *Conf. Empir. Methods Nat. Lang. Process.*, *Association for Computational Linguistics*, *Copenhagen*, *Denmark*, 2017, *pp.* 521–526, https://doi.org/10.18653/v1/D17-1054.

[9] *H. Chen*, *M. Sun*, *C. Tu*, *Y. Lin*, *Z. Liu*, "*Neural Sentiment classification with user and product attention*," *in Proc.* 2016 *Conf. Empir. Methods Nat. Lang. Process.*, *Association for Computational Linguistics*, *Austin*, *Texas*, 2016, *pp.* 1650–1659, https://doi.org/10.18653/v1/D16-1171.

[10] *Z. Wu*, *X.-Y. Dai*, *C. Yin*, *S. Huang*, *J. Chen*, "*Improving review representations with user attention and product attention for sentiment classification*," *in Thirty-Second AAAI Conf. Artif. Intell. AAAI*-18, 2018, *pp.* 5989–5996.

[11] *A. Kumar J*, *T.E. Trueman*, *E. Cambria*, "*A convolutional stacked bidirectional LSTM with a multiplicative attention mechanism for aspect category and sentiment detection*," *Cogn. Comput.* (2021), https://doi.org/10.1007/s12559-021-09948-0.

[12] *B. Liang*, *H. Su*, *L. Gui*, *E. Cambria*, *R. Xu*, "*Aspect-based sentiment analysis via affective knowledge enhanced graph convolutional networks*," *Knowl. Based Syst* 235 (2022), 107643, https://doi.org/10.1016/j.knosys.2021.107643.

[13] *W. Li*, *L. Zhu*, *Y. Shi*, *K. Guo*, *E. Cambria*, "*User reviews*: *sentiment analysis using lexicon integrated two-channel CNN–LSTM family models*," *Appl. Soft Comput.* 94 (2020), 106435, https://doi.org/10.1016/j.asoc.2020.106435.
[14] *W. Xue*, *T. Li*, "*Aspect based sentiment analysis with gated convolutional networks*," *in Proc. 56th Annu. Meet. Assoc. Comput. Linguist. Vol. 1 Long Pap.*, *Association for Computational Linguistics*, *Melbourne*, *Australia*, 2018, *pp.* 2514–2523, https://doi.org/10.18653/v1/P18-1234.
[15] *M.E. Basiri*, *S. Nemati*, *M. Abdar*, *E. Cambria*, "*ABCDM*: *an attention-based bidirectional CNN-RNN deep model for sentiment analysis*," *Future Gener. Comput. Syst.* 115 (2021) 279–294, https://doi.org/10.1016/j.future.2020.08.005.
[16] *M.S. Akhtar*, *A. Ekbal*, *E. Cambria*, "*How intense are you*? *Predicting intensities of emotions and sentiments using stacked ensemble*," *IEEE Comput. Intell. Mag.* 15 (2020) 64–75, https://doi.org/10.1109/MCI.2019.2954667.
[17] *T. Miyato*, *A.M. Dai*, *I. Goodfellow*, "*Adversarial training methods for semi-supervised text classification*," *in Int. Conf. Learn. Represent*, 2017.
[18] *C. Busso*, *S. Lee*, *S. Narayanan*, "*Analysis of emotionally salient aspects of fundamental frequency for emotion detection*," *IEEE Trans. Audio Speech Lang. Process.* 17 (2009) 582–596, https://doi.org/10.1109/TASL.2008.2009578.
[19] *D. Ghimire*, *J. Lee*, "*Geometric feature-based facial expression recognition in image sequences using multi-class adaboost and support vector machines*," *Sensors* 13 (2013) 7714–7734, https://doi.org/10.3390/s130607714.
[20] *A.A.S.Gunawan Sujono*, "*Face expression detection on Kinect using active appearance model and fuzzy logic*," *Procedia Comput. Sci.* 59 (2015) 268–274, https://doi.org/10.1016/j.procs.2015.07.558.
[21] *Y. Yan*, *Z. Zhang*, *S. Chen*, *H. Wang*, "*Low-resolution facial expression recognition*: *a filter learning perspective*," *Signal Process.* 169 (2020), 107370, https://doi.org/10.1016/j.sigpro.2019.107370.
[22] *Y. Yao*, *D. Huang*, *X. Yang*, *Y. Wang*, *L. Chen*, "*Texture and geometry scattering representation-based facial expression recognition in 2D+3D videos*," *ACM Trans. Multimed. Comput. Commun. Appl.* 14 (2018) 18.1–18.23, https://doi.org/10.1145/3131345.
[23] *Y. Zong*, *X. Huang*, *W. Zheng*, *Z. Cui*, *G. Zhao*, "*Learning from hierarchical spatiotemporal descriptors for micro-expression recognition*," *IEEE Trans. Multimed.* 20 (2018) 3160–3172, https://doi.org/10.1109/TMM.2018.2820321.
[24] *B. Sun*, *S. Cao*, *D. Li*, *J. He*, *L. Yu*, "*Dynamic Micro-expression recognition using knowledge distillation*," *IEEE Trans. Affect. Comput.* (2020), https://doi.org/10.1109/TAFFC.2020.2986962, 1–1.
[25] *K. Zhang*, *Y. Huang*, *Y. Du*, *L. Wang*, "*Facial expression recognition based on deep evolutional spatial-temporal networks*," *IEEE Trans. Image Process* 26 (2017) 4193–4203, https://doi.org/10.1109/TIP.2017.2689999.
[26] *F. Zhang*, *T. Zhang*, *Q. Mao*, *C. Xu*, "*Joint pose and expression modeling for facial expression recognition*," *in Proc.* 2018 *IEEECVF Conf. Comput. Vis. Pattern Recognit.*, *IEEE*, *Salt Lake City*, *UT*, *USA*, 2018, *pp.* 3359–3368, https://doi.org/10.1109/CVPR.2018.00354.

[27] *F. Zhang*, *T. Zhang*, *Q. Mao*, *C. Xu*, "*Geometry guided pose-invariant facial expression recognition*," *IEEE Trans. Image Process*. 29 (2020) 4445–4460, *https://doi.org/*10.1109/*TIP*.2020.2972114.
[28] *M. Wollmer*, *F. Weninger*, *T. Knaup*, *B. Schuller*, *C. Sun*, *K. Sagae*, *L.-P. Morency*, "*YouTube movie reviews*: *sentiment analysis in an audio-visual context*," *IEEE Intell. Syst*. 28 (2013) 46–53, *https://doi.org/*10.1109/*MIS*.2013.34.
[29] *Y. Zhang*, *Z.-R. Wang*, *J. Du*, "*Deep fusion*: *an attention-guided factorized bilinear pooling for audio-video emotion recognition*," *in Int. Jt. Conf. Neural Netw*, 2019, *https://doi.org/*10.1109/*IJCNN*.2019.8851942.
[30] *S. Zhao*, *Y. Ma*, *Y. Gu*, *J. Yang*, *T. Xing*, *P. Xu*, *R. Hu*, *K. Keutzer*, "*An end-to-end visual-audio attention network for emotion recognition in user-generated videos*," *Proc. AAAI Conf. Artif. Intell*. 34 (2020) 303–311, *https://doi.org/*10.1609/*aaai.v*34*i*01.5364.
[31] *S. Zhang*, *S. Zhang*, *T. Huang*, *W. Gao*, *Q. Tian*, "*Learning affective features with a hybrid deep model for audio–visual emotion recognition*," *IEEE Trans. Circuits Syst. Video Technol*. 28 (2018) 3030–3043, *https://doi.org/*10.1109/*TCSVT*.2017.2719043.
[32] *M.S. Hossain*, *G. Muhammad*, "*Emotion recognition using deep learning approach from audio–visual emotional big data*," *Inf. Fusion*. 49 (2019) 69–78, *https://doi.org/*10.1016/*j.inffus*.2018.09.008.
[33] *Robinson*, *D. L.* (2009). "*Brain function*, *mental experience and personality*". *The Netherlands Journal of Psychology*. *pp*. 152–167.
[34] *Shaver*, *P.*; *Schwartz*, *J.*; *Kirson*, *D.*; *O'connor*, *C.* (1987). "*Emotion knowledge*: *further exploration of a prototype approach*". *Journal of Personality and Social Psychology*. 52 (6): 1061–86. *doi*:10.1037/0022-3514.52.6.1061. *PMID* 3598857.
[35] *Parrott*, *W.* (2001). *Emotions in Social Psychology*. *Key Readings in Social Psychology*. *Philadelphia*: *Psychology Press*. *ISBN* 978-0863776830.
[36] "*Basic Emotions—Plutchik*". *Personalityresearch.org*. *Retrieved* 1 *September* 2017.
[37] *Plutchik*, *R.* "*The Nature of Emotions*". *American Scientist*. *Archived from the original on July* 16, 2001. *Retrieved* 14 *April* 2011.
[38] "*Robert Plutchik's Psychoevolutionary Theory of Basic Emotions*". *Adliterate.com*. *Retrieved* 2017-06-05.
[39] *Jonathan Turner* (1 *June* 2000). *On the Origins of Human Emotions*: *A Sociological Inquiry Into the Evolution of Human Affect*. *Stanford University Press*. *p*. 76. *ISBN* 978-0-8047-6436-0.
[40] *Atifa Athar*; *M. Saleem Khan*; *Khalil Ahmed*; *Aiesha Ahmed*; *Nida Anwar* (*June* 2011). "*A Fuzzy Inference System for Synergy Estimation of Simultaneous Emotion Dynamics in Agents*". *International Journal of Scientific & Engineering Research*. 2 (6).
[41] *TenHouten*, *Warren D.* (1 *December* 2016). *Alienation and Affect*. *Taylor & Francis*. *ISBN* 9781317678533. *Retrieved* 25 *June* 2019 – *via Google Books*.
[42] *Chorianopoulos*, *Konstantinos*; *Divitini*, *Monica*; *Hauge*, *Jannicke Baalsrud*; *Jaccheri*, *Letizia*; *Malaka*, *Rainer* (24 *September* 2015). *Entertainment Computing - ICEC* 2015: 14*th International Conference*, *ICEC* 2015, *Trondheim*, *Norway*, *September* 29 - *October* 2, 2015, *Proceedings*. *Springer*. *ISBN* 978331924
[43] *Plutchik*, *Robert* (25 *June* 1991). *The Emotions*. *University Press of America*. *ISBN* 9780819182869. *Retrieved* 25 *June* 2019 – *via Google Books*.

[44] *O'Shaughnessy*, *John* (4 *December* 2012). *Consumer Behaviour*: *Perspectives*, *Findings and Explanations*. *Macmillan International Higher Education*. *ISBN* 9781137003782. *Retrieved* 25 *June* 2019 – *via Google Books*.
[45] *Kort*, *B*.; *Reilly*, *R*.; *Picard*, *R.W*. (2001). "*An affective model of interplay between emotions and learning*: *Reengineering educational pedagogy-building a learning companion*". *Proceedings IEEE International Conference on Advanced Learning Technologies*. *pp*. 43–46. *doi*:10.1109/*ICALT*.2001.943850. *ISBN* 0-7695-1013-2. *S*2*CID* 9573470 – *via www.academia.edu*.
[46] *Lim*, *Lucinda*, *et al*. "*Where is the emotion*? *Dissecting a multi-gap network for image emotion classification*." 2020 *IEEE International Conference on Image Processing* (*ICIP*). *IEEE*, 2020.

## *Reference URL*


[1] *Late Fusion-Late Fusion*. (2020). *Retrieved from* https://blog.csdn.net/qq_40438165/article/details/106358697

[2] 多模态中的 *Late Fusion*. *Retrieved from* https://zhuanlan.zhihu.com/p/664065153

[3] *MultiBench*: *Multiscale Benchmarks for Multimodal Representation Learning*. *Retrieved from* https://github.com/pliang279/MultiBench

[4] 情绪的详细分类合集. *Retrieved from* https://zhuanlan.zhihu.com/p/544848260

[5] *Wikipedia*. (*Year*). *Emotion Classification*. *Retrieved from* https://en.wikipedia.org/wiki/Emotion_classification#Six_emotion_axes

[6] *The Best And Worst Things Each Emotion Did In Inside Out*. *Retrieved from* https://thiepnhanai.com/the-best-and-worst-things-each-emotion-did-in-inside-out/

[7] 情感计算研究进展、现状及趋势. *Retrieved from* http://hlt.hitsz.edu.cn/info/1001/1402.htm

[8] 情感计算工作小结. *Retrieved from* https://zhuanlan.zhihu.com/p/83006064

[9] 多模态情感分类简介. *Retrieved from* https://zhuanlan.zhihu.com/p/620490187

[10] 情感建模. *Retrieved from* https://www.jianshu.com/p/5403d41e7b31

[11] 情感计算：让机器更加智能. *Retrieved from* https://zhuanlan.zhihu.com/p/86073203

[12] 注意力机制(*Attention Mechanism*)浅谈. *Retrieved from* https://zhuanlan.zhihu.com/p/364819787

[13] 多模态机器学习. *Retrieved from* https://zhuanlan.zhihu.com/p/431286411

[14] 多模态融合 *fusion* 的各种操作. *Retrieved from* https://zhuanlan.zhihu.com/p/152234745

[15] 情感计算综述. *Retrieved from* https://blog.csdn.net/ChristopherI/article/details/130849281