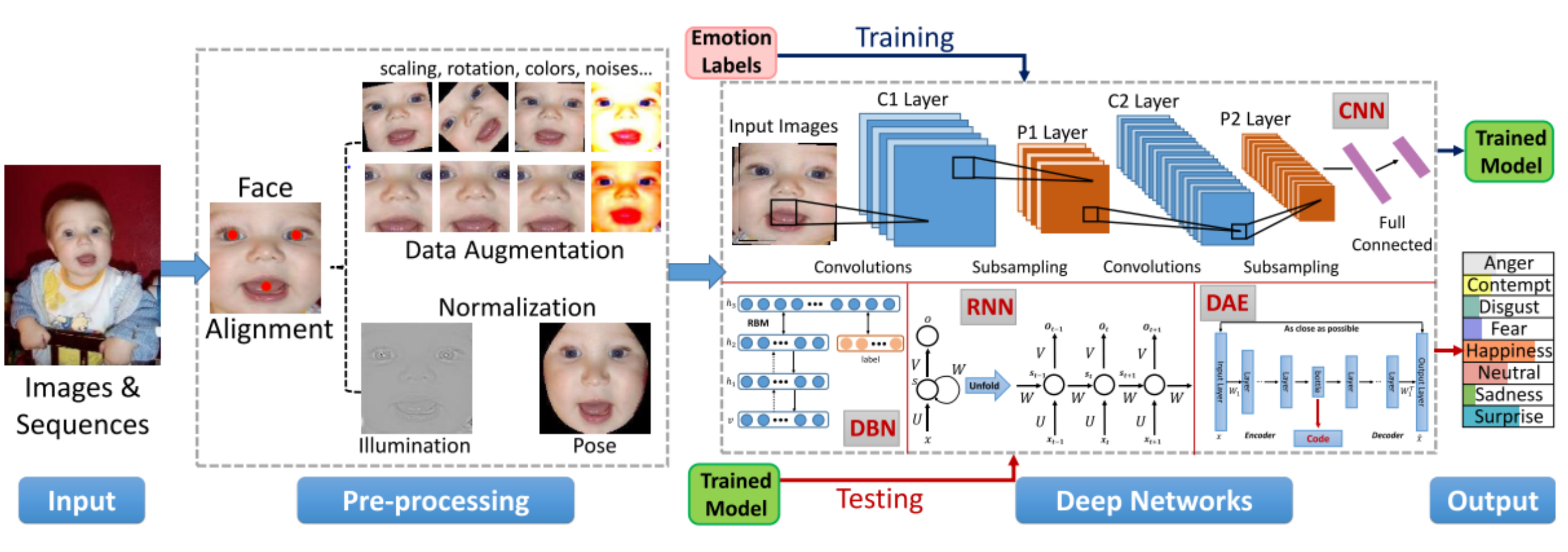
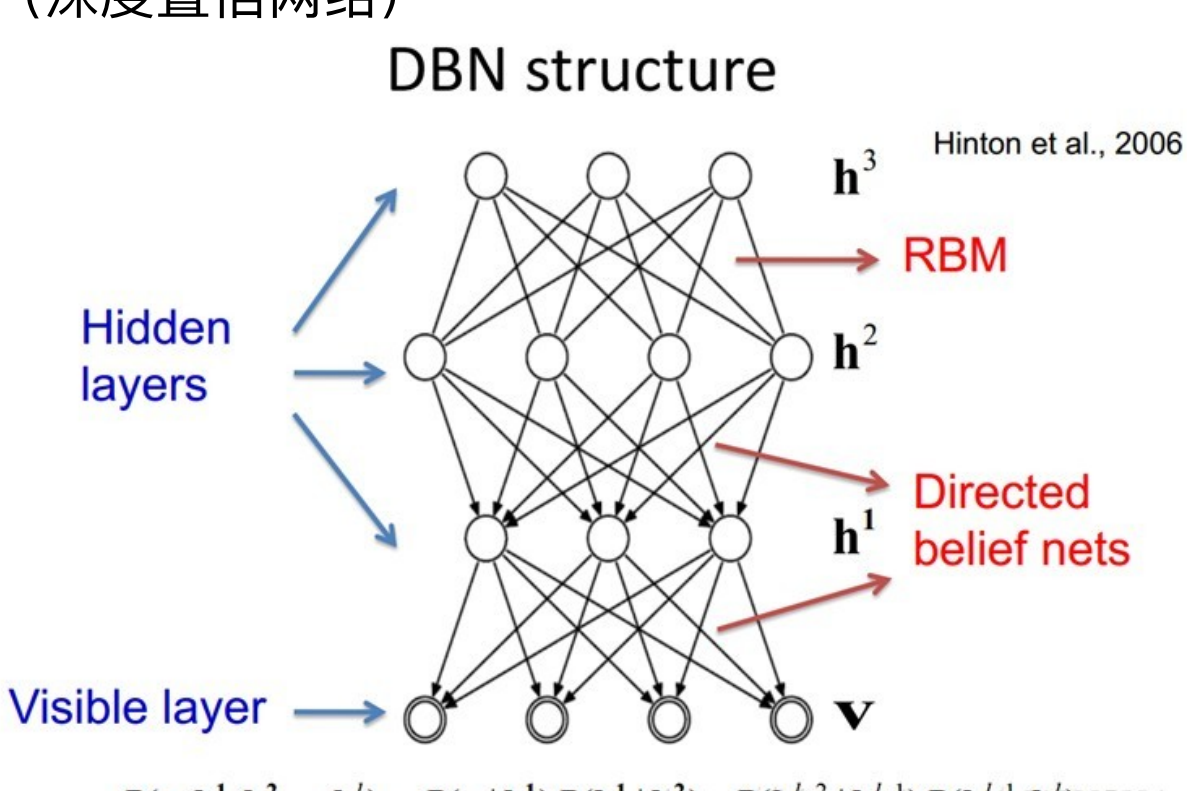
- ## Introduction
  - 人脸表情是最直接、最有效的情感识别模式。
  - 早在20世纪Ekman等专家通过跨文化调研提出7类基本表情：
    - 生气、害怕、厌恶、开心、悲伤、惊讶及中性。
  - 之后在2014年在PNAS又有提出复合表情概念，并指出多种离散的基础表情进行结合形成复合表情。
    - 惊喜、悲愤。。。
  - 这篇文章为一篇人脸表情方面的综述，主要针对的是基于深度学习方面的人脸表情研究，并对现有方法进行总结和整理，包括步骤、数据集以及网络设计技巧。
    - 表情数据集由实验室小样本数据转向现实生活多样化大规模
    - 算法由传统特征转向深度特征

- ## Deep Facial Expression Recognition

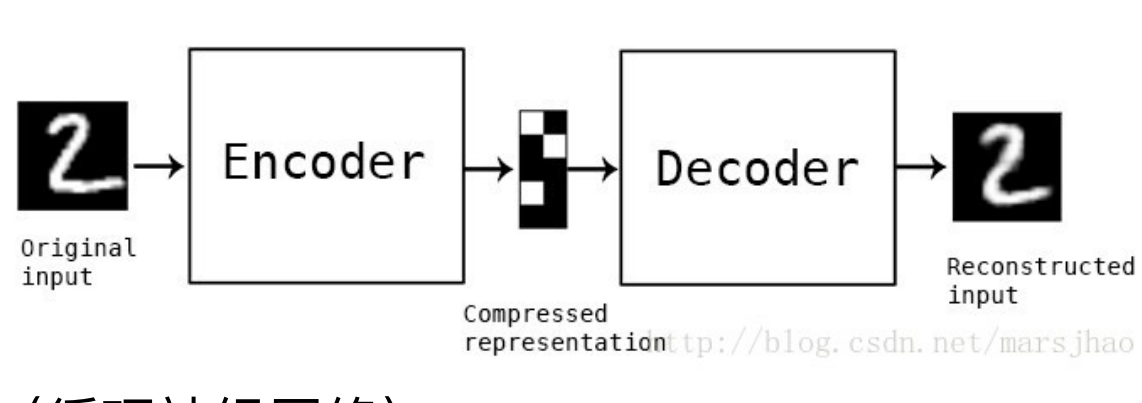


- Fig. 2. The general pipeline of deep facial expression recognition systems.
- 预处理
  - Face alignment
    - 人脸检测：需要检测人脸位置，分离非人脸部分
      - opencv及Matalb中的自带人脸检测算法（Haarcascades）
      - MTCNN
      - SSD（目前在用）
    - 人脸对齐：矫正人脸，减少人脸旋转角度，方便统一模板
      - IntraFace
      - MoT
      - Dlib
      - DRMF
    - 多模型特征结合的方法
  - 数据增强
    - 离线数据增强
      - 旋转，平移、缩放、噪声、亮度、GAN。。。
    - 动态数据增强
      - 在输入网络时随机裁剪、翻转
  - 人脸归一化
    - 光照归一化（处理光照问题）
      - INFace toolbox（Matlab）
      - 直方图均衡化（常用且最为可靠）
    - 姿态归一化（获得正面照）
      - 3D纹理模型
      - GAN
- 深度网络
  - CNN（卷积神经网络）
    - R-CNN、Faster R-CNN等检测算法用于FER（【35】【80】-【87】）
  - DBN（深度置信网络）

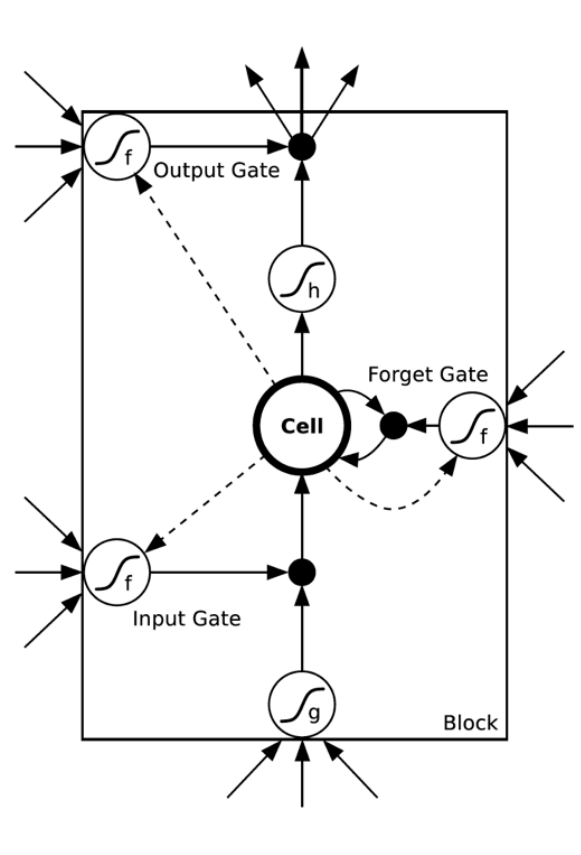


$$P(\mathbf{v},\mathbf{h}^1,\mathbf{h}^2,...,\mathbf{h}^l) = P(\mathbf{v}\,|\,\mathbf{h}^1)P(\mathbf{h}^1\,|\,\mathbf{h}^2)...P(\mathbf{h}^{l-2}\,|\,\mathbf{h}^{l-1})P(\mathbf{h}^{l-1},\mathbf{h}^l)$$

  - DAE（深度自编码器）



  - RNN（循环神经网络）
    - LSTM



- 分类器
  - 网络输出直接分类
  - 结合SVM或随机森林

- ## Faial Expression DataBases
  - CK+：包含123个主题的593个视频序列，并有327个序列标记了7中基本表情
  - MMI：包含32个主题326个序列，213个序列标记了6种基本表情（没有轻视）
  - JAFFE：包含10位日本女性的表情样本，每人有3-4张图片，6种表情标签
  - TFD：包含112,234张图片，其中4,178张标注了7种基本表情并统一到48*48大小
  - FER2013：包含28,709张训练集，3,589张验证集，3,589张测试集，6种表情标签并统一到48*48大小
  - AFEW：用于EmotiW评估平台，7种表情标签
  - SFEW：AFEW的静态图版本
  - EmotioNet：AU数据集，950,000自动标注，25,000手动标注，11种AUs，6种表情标签，10种复杂标签
  - AffectNet：目前最大的标签数据库，450,000张手动标注数据

Database	Samples	Subjects	Condit.	Elicit.	Expression distribution	Access
CK+ [102]	593 image sequences	123	Lab	P & S	6 basic expressions plus contempt and neutral	http://www.consortium.ri.cmu.edu/ckagree/
MMI [103], [104]	740 images and 2,900 videos	25	Lab	P	6 basic expressions plus neutral	https://mmifacedb.eu/
JAFFE [105]	213 images	10	Lab	P	6 basic expressions plus neutral	http://www.kasrl.org/jaffe.html
TFD [106]	112,234 images	N/A	Lab	P	6 basic expressions plus neutral	josh@mplab.ucsd.edu
FER-2013 [10]	35,887 images	N/A	Web	P & S	6 basic expressions plus neutral	https://www.kaggle.com/c/challenges-in-representation-learning-facial-expression-recognition-challenge
AFEW 7.0 [15]	1,809 videos	N/A	Movie	P & S	6 basic expressions plus neutral	https://sites.google.com/site/emotiwchallenge/
SFEW 2.0 [13]	1,766 images	N/A	Movie	P & S	6 basic expressions plus neutral	https://cs.anu.edu.au/few/emotiw2015.html
Multi-PIE [107]	755,370 images	337	Lab	P	Smile, surprised, squint, disgust, scream and neutral	http://www.flintbox.com/public/project/4742/
BU-3DFE [108]	2,500 images	100	Lab	P	6 basic expressions plus neutral	http://www.cs.binghamton.edu/~lijun/Research/3DFE/3DFE_Analysis.html
Oulu-CASIA [109]	2,880 image sequences	80	Lab	P	6 basic expressions	http://www.cse.oulu.fi/CMV/Downloads/Oulu-CASIA
RaFD [110]	1,608 images	67	Lab	P	6 basic expressions plus contempt and neutral	http://www.socsci.ru.nl:8180/RaFD2/RaFD
KDEF [111]	4900 images	70	Lab	P	6 basic expressions plus neutral	http://www.emotionlab.se/kdef/
EmotioNet [112]	1,000,000 images	N/A	Web	P & S	23 basic expressions or compound expressions	http://cbcsl.ece.ohio-state.edu/dbform_emotionet.html
RAF-DB [113]	29672 images	N/A	Web	P & S	6 basic expressions plus neutral and 12 compound expressions	http://www.whdeng.cn/RAF/model1.html
AffectNet [114]	450,000 images (labeled)	N/A	Web	P & S	6 basic expressions plus neutral	http://mohammadmahoor.com/databases-codes/

† The number of samples and number of subjects for each dataset are taken from the reference paper and may be different from the actual data that are available for basic emotion recognition. See text for details.

- ## The State Of The Art
  - 针对静态图片的深度FER网络
    - Pre-training and fine-tuning
      - 提前在人脸识别数据集（CASIA WebFace，CFW等）上进行预训练之后再FER数据集上进行fine-tuning效果会有所提升
      - 多级fine-tuning（采用不同的微调组合进行fine-tuning）：

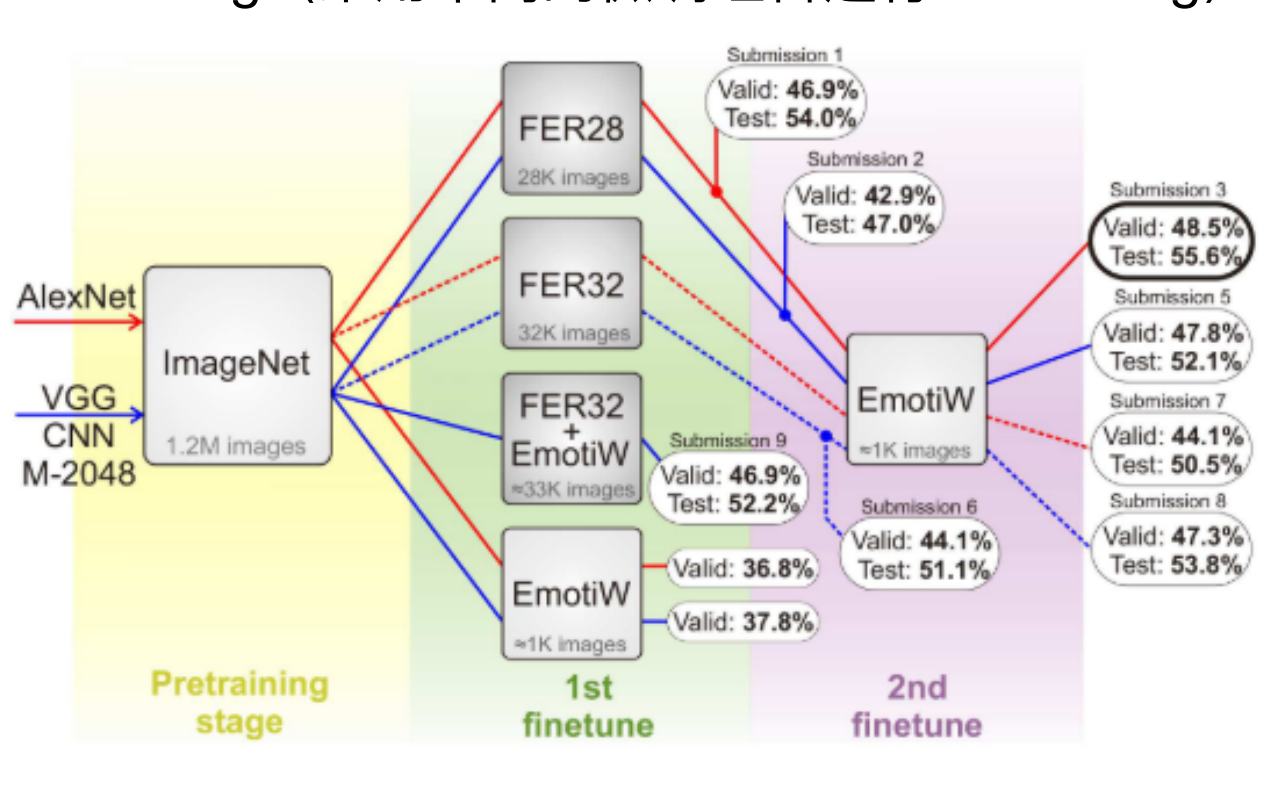


Fig. 3. Flowchart of the different fine-tuning combinations used in [29]. Here, “FER28” and “FER32” indicate different parts of the FER2013 datasets. “EmotiW” is the target dataset. The proposed two-stage fine-tuning strategy (Submission 3) exhibited the best performance.

      - 人脸识别数据集与FER数据集之间差距较大，有可能人脸占主体，【31】提出一种FaceNet2ExpNet方式。（a阶段通过分布函数及正则化将网络向Face Net推进，B阶段增加随机初始化全连接层，进行联合训练）（待查？）

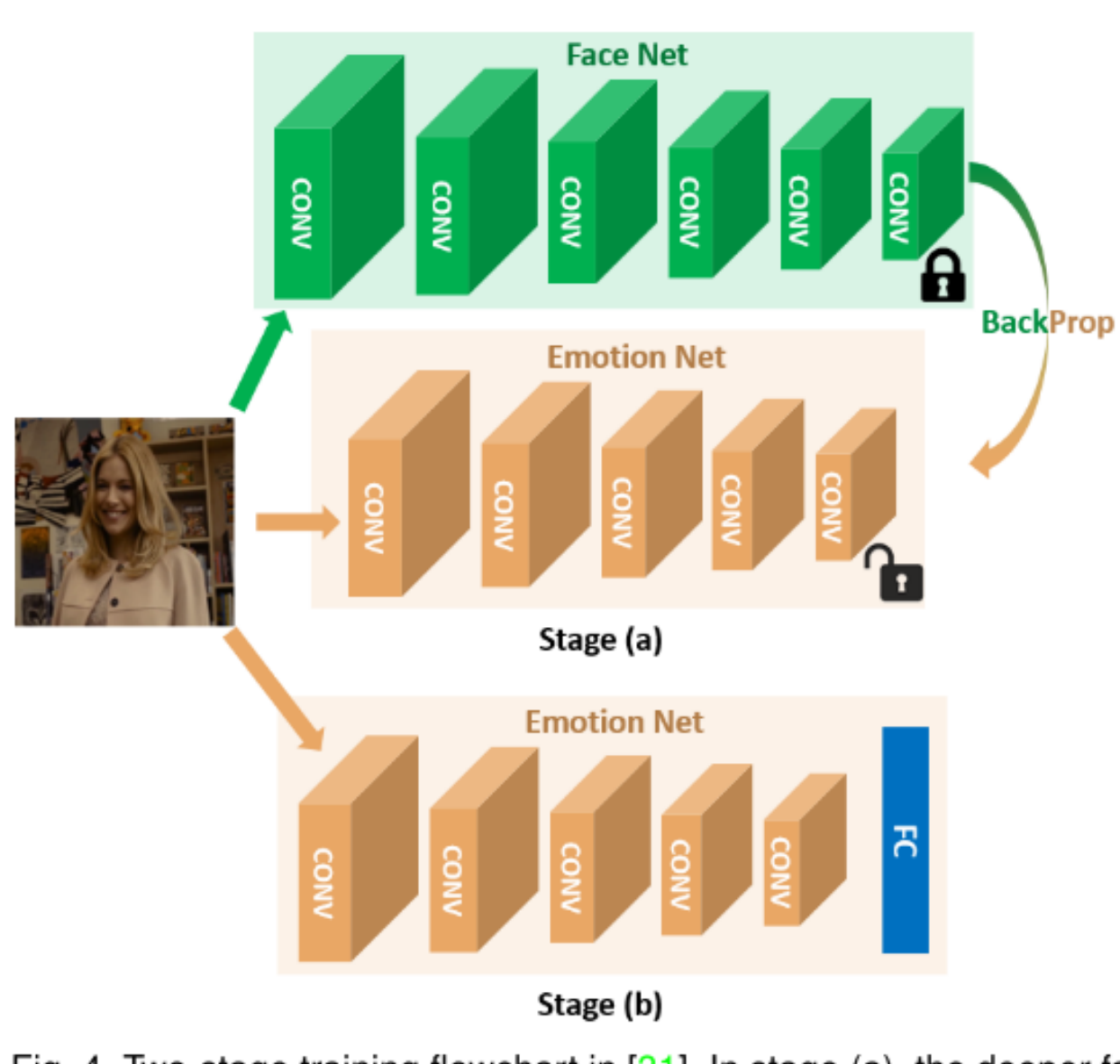
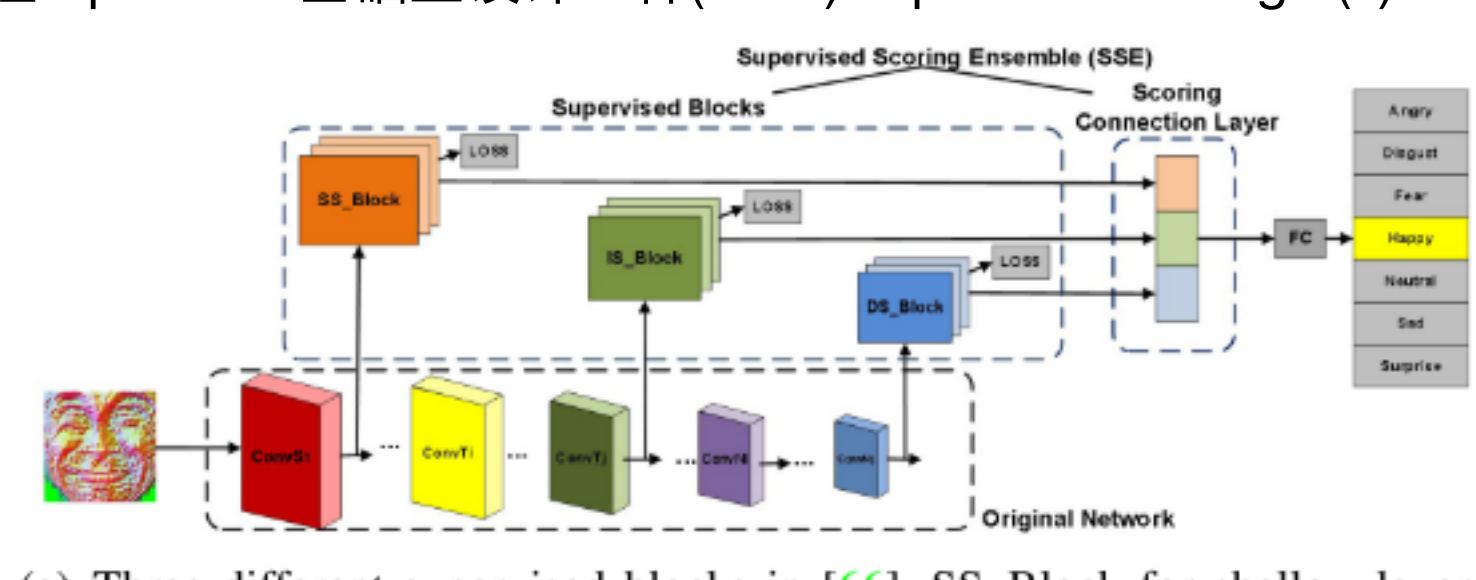
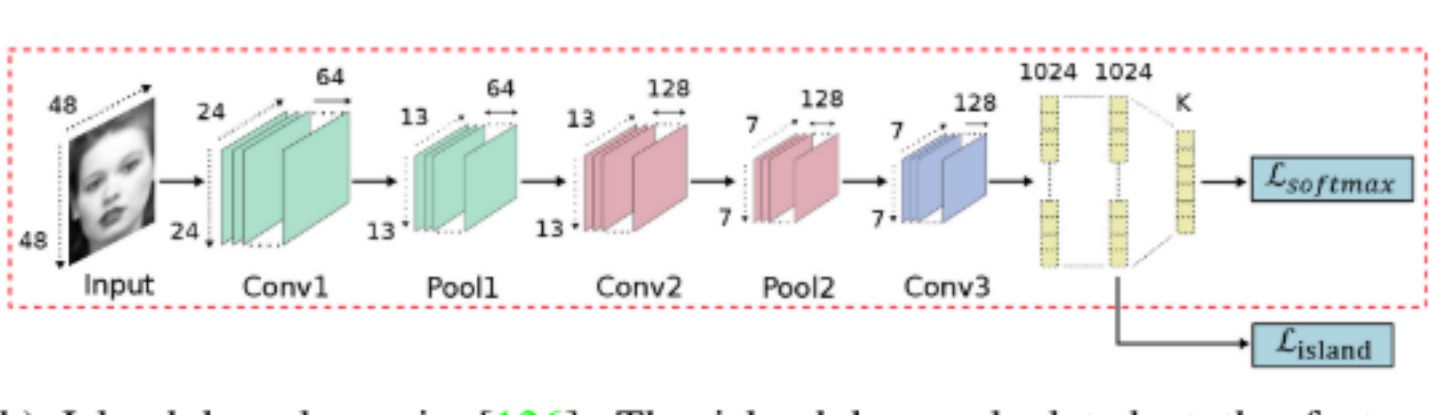


Fig. 4. Two-stage training flowchart in [31]. In stage (a), the deeper face net is frozen and provides the feature-level regularization that pushes the convolutional features of the expression net to be close to the face net by using the proposed distribution function. Then, in stage (b), to further improve the discriminativeness of the learned features, randomly initialized fully connected layers are added and jointly trained with the whole expression net using the expression label information.

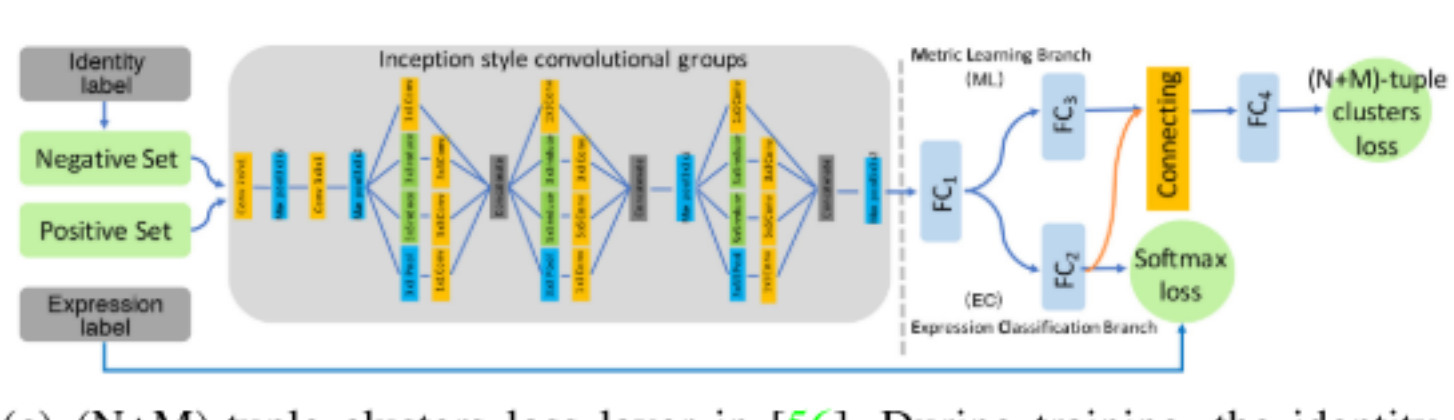
    - Diverse network input
      - 采用一些传统特征作为网络输入，以解决RGB图像信息不明确问题
        - LBP
        - SIFT
        - HOG
        - 组合特征
    - Auxiliary blocks & layers
      - HoloNet：CReLU+残差结构+inception块
      - 增加不同层次监督进行分级聚合---Fig.6(a)
      - 设计island损失函数以针对类间距离---Fig.6(b)
      - 在triplet loss基础上设计一种(N+M)-tuples loss ---Fig.6(c)



(a) Three different supervised blocks in [66]. SS_Block for shallow-layer supervision, IS_Block for intermediate-layer supervision, and DS_Block for deep-layer supervision.



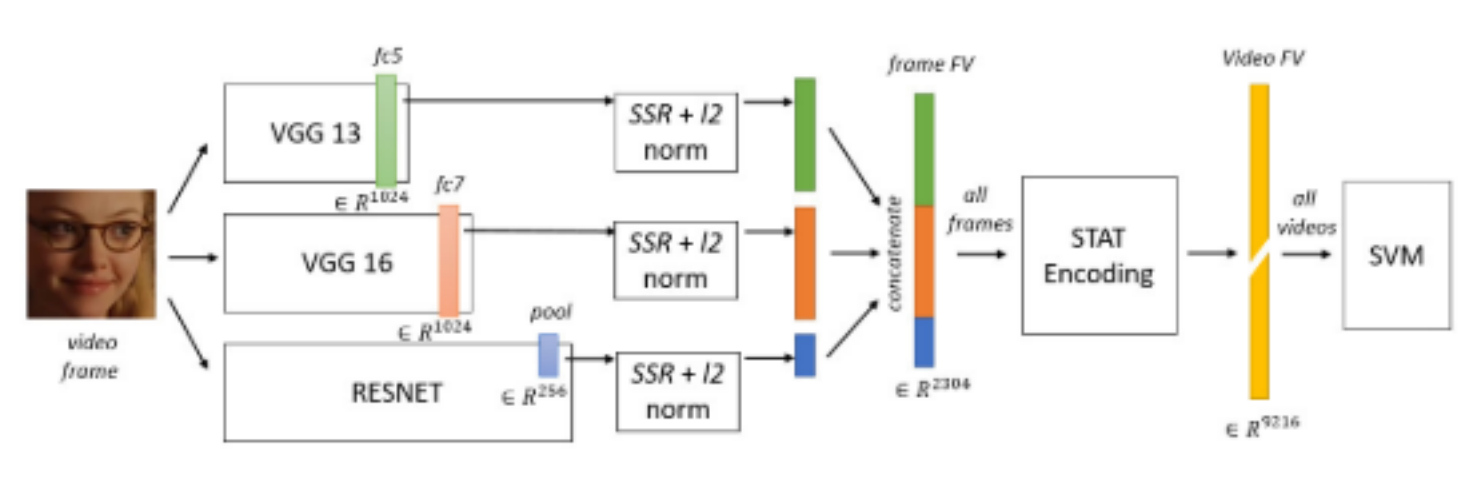
(b) Island loss layer in [126]. The island loss calculated at the feature extraction layer and the softmax loss calculated at the decision layer are combined to supervise the CNN training.



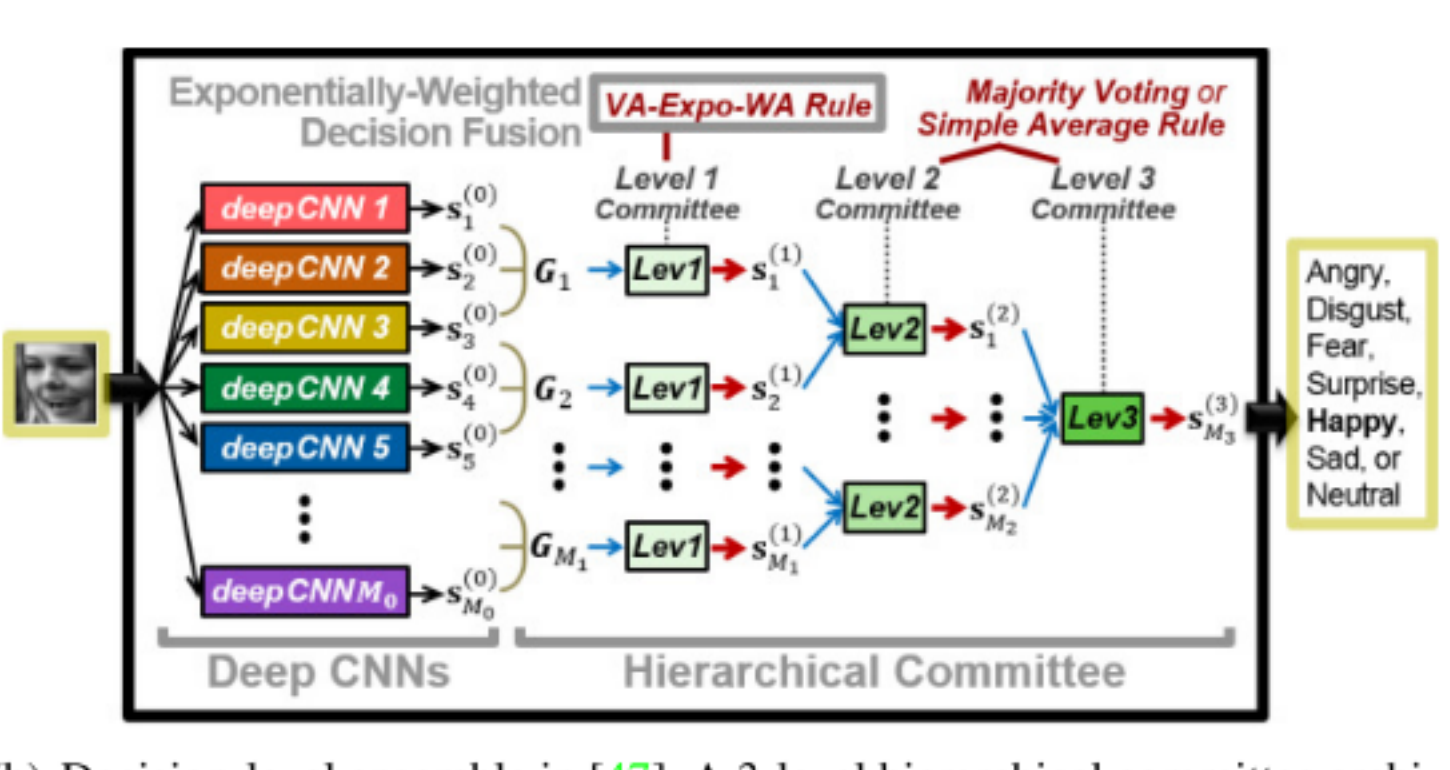
(c) (N+M)-tuple clusters loss layer in [56]. During training, the identity-aware hard-negative mining and online positive mining schemes are used to decrease the inter-identity variation in the same expression class.

Fig. 6. Representative functional layers or blocks that are specifically designed for deep facial expression recognition.

    - Network ensemble
      - 联合多个网络的效果通常不单个网络要好
      - 应用不同网络进行学习，通过串联以获得一个完整图像图像特征进行分类---Fig.7(a)
      - 在决策方面进行优化，例如Majority Voting、Simple Average、Weighter Average---Fig.7(b)



(a) Feature-level ensemble in [60]. Three different features (*fc5* of VGG13 + *fc7* of VGG16 + *pool* of Resnet) after normalization are concatenated to create a single feature vector (FV) that describes the input frame.



(b) Decision-level ensemble in [47]. A 3-level hierarchical committee architecture with hybrid decision-level fusions was proposed to obtain sufficient decision diversity.

Fig. 7. Representative network ensemble systems at the feature level and decision level.

    - Multitask networks
      - 可以引入transfer knowledge将头部姿势、关键点、识别等影响因素一起作为网络任务，以提高稳定性。
      - 采用不同损失函数将人脸识别与FER任务相结合，一同训练网络（MSCNN）---Fig.8

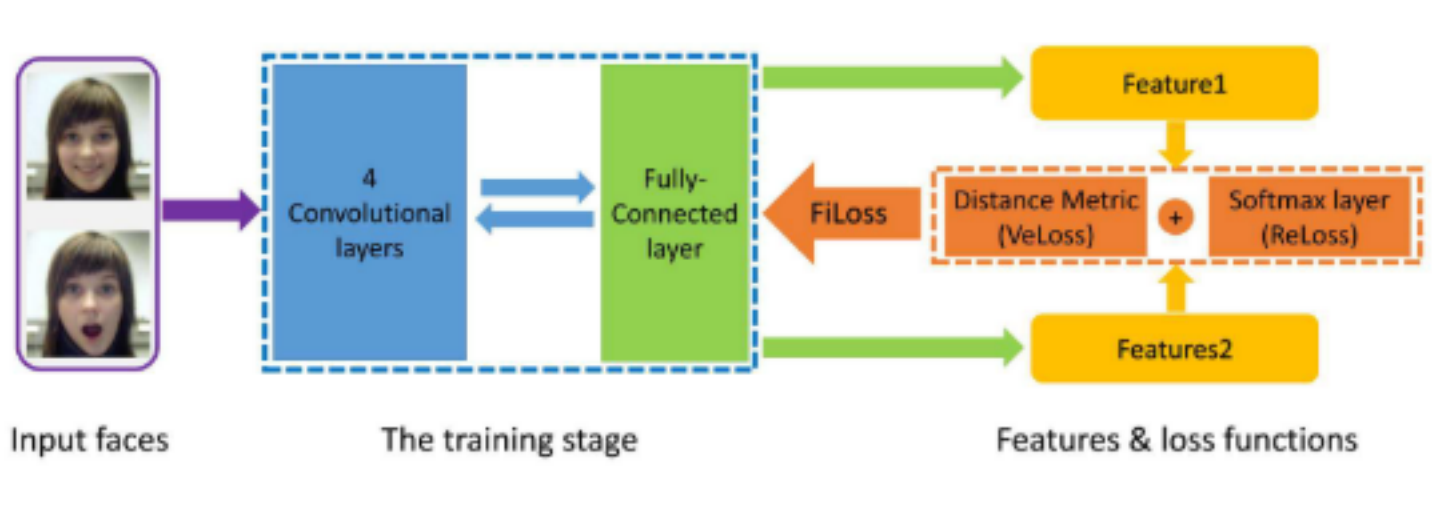


Fig. 8. Representative multitask network for FER. In the proposed MSCNN [165], a pair of images is sent into the MSCNN during training. The expression recognition task with cross-entropy loss, which learns features with large between-expression variation, and the face verification task with contrastive loss, which reduces the variation in within-expression features, are combined to train the MSCNN.

    - Cascaded networks
      - 级联网络能够将不同或相同任务模块进行组合，以前一个网络输出作为后一个网络输入的形式，逐步将无关因子过滤

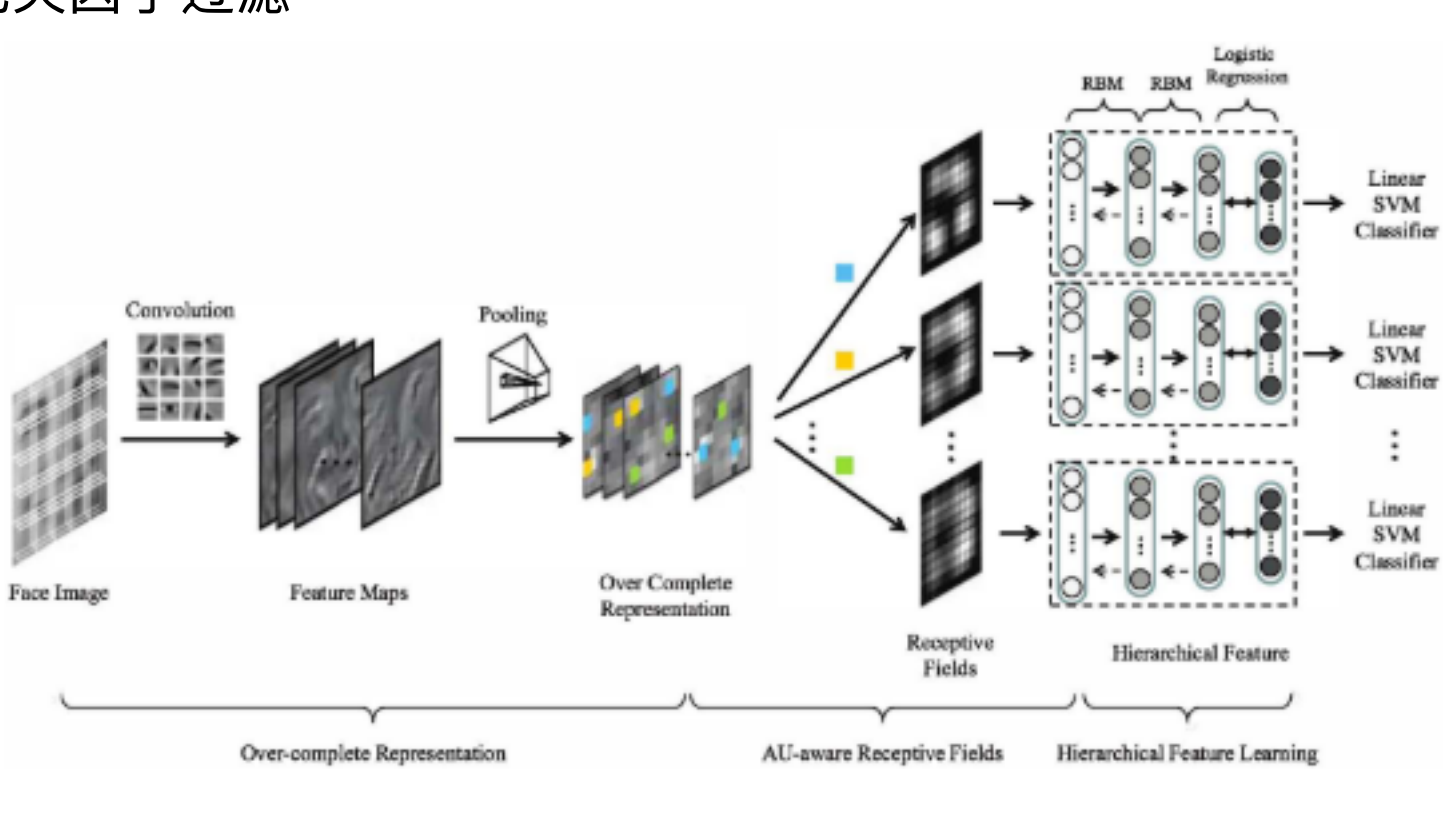


Fig. 9. Representative cascaded network for FER. The proposed AU-aware deep network (AUDN) [122] is composed of three sequential modules: in the first module, a 2-layer CNN is trained to generate an over-complete representation encoding all expression-specific appearance variations over all possible locations; in the second module, an AU-aware receptive field layer is designed to search subsets of the over-complete representation; in the last module, a multilayer RBM is exploited to learn hierarchical features.

  - 针对动态图片序列的深度FER网络
    - Frame aggregation
      - 将多帧图片进行聚合往往能获得一个比较稳定的结果输出
      - 从决策上：连接连续的n个特征向量作为后续划分标准
      - 从特征上：在特征向量中应用特征模型（线性模型、高斯分布模型等）
    - Expression Intensity network
      - 引入表情强度的相关信息，学习不同表情强度之间的相同信息
    - Deep spatio-temporal FER network
      - 引入时空信息，通常将图片序列进行训练后，利用RNN网络学习特征信息

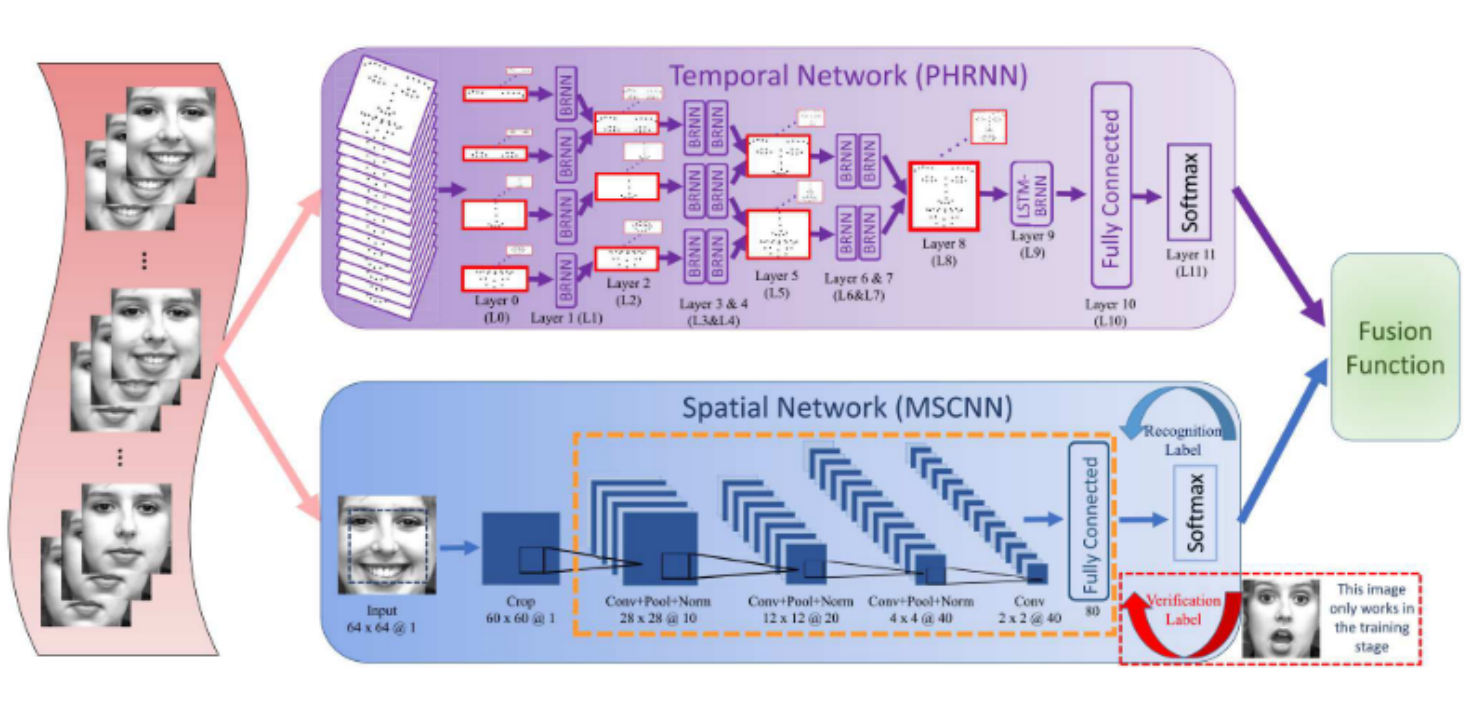


Fig. 13. The spatio-temporal network proposed in [165]. The temporal network PHRNN for landmark trajectory and the spatial network MSCNN for identity-invariant features are trained separately. Then, the predicted probabilities from the two networks are fused together for spatio-temporal FER.

- ## 几点思考：
  - 数据集的储备（小规模数据集，跨数据集准确度下降等问题）
  - 结合语音、三维图像等信息进行情感识别
  - 图片特征的鲁棒性研究
  - 人脸表情关键因素的研究
  - 复合表情方面的多种挑战