第 45 卷 第 5 期 计 算 机 工 程 2019 年 5 月
Vol. 45 No. 5 Computer Engineering May 2019

·多媒体技术及应用· 文章编号:1000-3428(2019)05-0256-05 文献标志码:A 中图分类号:TP391.4

基于微调策略的多线索融合人脸活体检测

胡 斐[1a],文 畅[1a],谢 凯[1b],贺建飚[2]

(1. 长江大学 a. 计算机科学学院; b. 电子信息学院,湖北 荆州 434023;
2. 中南大学 信息科学与工程学院,长沙 410083)

摘 要: 为解决身份认证过程中可能会出现的打印攻击、视频重播攻击等安全问题,提出一种多线索融合人脸活体检测方法。利用金字塔 LK 光流追踪视频帧并将其进行剪切波变换,以获取图像质量特征,通过卷积神经网络对数据集进行网络微调,得到真假活体。在 Print-attack 数据库和 CISIA 数据库上进行实验,结果表明,与 LFDNet 方法相比,该方法具有较高的人脸活体检测准确率,可用于抵制欺骗攻击。

关键词: 卷积神经网络; 微调策略; 剪切波变换; 金字塔 LK 光流; 人脸活体检测

中文引用格式: 胡斐,文畅,谢凯,等. 基于微调策略的多线索融合人脸活体检测[J]. 计算机工程,2019,45(5):256-260.
英文引用格式: HU Fei, WEN Chang, XIE Kai, et al. Multi-cue fusion face liveness detection based on fine-tuning strategy[J]. Computer Engineering, 2019, 45(5): 256-260.

Multi-cue Fusion Face Liveness Detection Based on Fine-tuning Strategy

HU Fei[1a], WEN Chang[1a], XIE Kai[1b], HE Jianbiao[2]

(1a. School of Computer Science; 1b. School of Electronic Information, Yangtze University, Jingzhou, Hubei 434023, China;
2. College of Information Science and Engineering, Central South University, Changsha 410083, China)

【Abstract】 To solve the security problems such as print attacks and video replay attacks that may occur during the authentication process, a multi-cue fusion method for detecting human faces is proposed. The pyramid Lucas-Kanade (LK) optical flow is used to track the video frame, and it is subjected to shear wave transform to obtain image quality features. The Convolutional Neural Network (CNN) is used to fine tune the data set to obtain true and false liveness bodies. Experiments on the Print-attack database and the CISIA database show that compared with the LFDNet method, this method has higher face liveness detection accuracy and can be applied to spoofing attacks resistance.

【Key words】 Convolutional Neural Network(CNN); fine-tuning strategy; shear wave transform; pyramid Lucas-Kanade(LK) optical flow; face liveness detection

DOI: 10.19678/j.issn.1000-3428.0050721

0 概述

基于生物特征(尤其是人脸)的身份识别已被广泛应用于信息安全与身份认证领域,而各种欺骗攻击手段也接踵而至,企图破解各种认证方式。为能够检测出是否为本人操作获取相关权限并保障个人的信息安全,开始对活体与欺骗攻击进行区别,即活体检测。

活体检测方法主要分为 3 类:运动信息分析,纹理信息分析和人脸三维结构分析。运动信息分析如分析人脸的眨眼[1]、张嘴[2]、点头[3]等动态特征,文献[4]利用多层卷积神经网络(Convolutional Neural Network,CNN)提取眨眼信息进行活体检测。文献[5]采用多线索融合处理空间域和时间域的活体检测。文献[6]利用前后背景的光流相关性来进行活体检测。上述方法虽然对照片攻击有较好的检测效果,但在面临高质量的视频攻击时则难以达到较高的精确度。纹理信息分析通过对三维活体人脸和照片成像后的差异性来提取纹理特征进行活体判断。文献[7]提出一种使用微纹理分析的方式进行活体检测,但在图像质量较差时效果不佳。文献[8]将纹理特征结合人脸结构特征,能够避免图像质量较差的缺陷,但算法复杂度较大,对时间要求高。人脸三维结构分析主要利用照片、屏幕作为二维平面,

基金项目: 国家自然科学基金(61272147)。
作者简介: 胡 斐(1996—),男,硕士研究生,主研方向为信号处理、图像与视频处理;文 畅(通信作者),讲师;谢 凯,教授;贺建飚,副教授。
收稿日期: 2018-03-12 **修回日期:** 2018-05-07 **E-mail:** wenchang2016paper@163.com

人脸作为三维立体来进行活体检测。文献[9]提出三维人脸和二维平面光的反射不同的思想。文献[10]根据不同材料(皮肤与非皮肤)光反射曲线的不同来进行活体检测,但需要额外的红外设备,不易实现。

基于以上研究,本文提出一种多线索融合人脸活体检测方法。该方法采用金字塔 LK 光流追踪视频序列的运动信息,利用剪切波变换[11]在多维数据中捕获异性特征,并提取纹理特征,通过卷积神经网络[12]进行微调,以提高检测的准确率。

1　多线索融合人脸活体检测

本文利用卷积神经网络提取视频序列中的运动信息以及图像质量信息,通过训练网络判断活体人脸。从视频样本中提取一个 10 帧的连续视频序列,将其利用金字塔 LK 光流追踪动态信息,同时对第 1 帧进行剪切波变换,最终把提取的光流特征和图像质量特征输入到卷积神经网络,再使用 Caffe 框架进行微调,使其具备判断是活体或非活体的能力。

1.1　金字塔 LK 光流

光流是图像表面运动的速度,由此推导出基于该模型的约束方程,算法流程如图 1 所示。

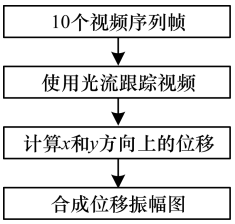

图 1　金字塔 LK 光流追踪算法流程

在 t 时刻,图像上的点 $\boldsymbol{P}=(x,y)$ 的灰度值是 $\boldsymbol{I}(x,y,t)$,在经过 $\mathrm{d}t$ 时间后,与之匹配的灰度值变为 $\boldsymbol{I}=(x+\mathrm{d}x,y+\mathrm{d}y,t+\mathrm{d}t)$。当 $\mathrm{d}t$ 趋近于 0 时,2 点的灰度值不变,如式(1)所示。

$$\boldsymbol{I}=(x+\mathrm{d}x,y+\mathrm{d}y,t+\mathrm{d}t)=\boldsymbol{I}(x,y,t) \tag{1}$$

当图像的灰度值随着 x、y、t 这 3 个变量变化时,式(1)由泰勒级数展开为:

$$I_x\mu+I_y\nu+I_t=0 \tag{2}$$

光流场基本方程的矢量形式可表示为:

$$\nabla\boldsymbol{I}\cdot\boldsymbol{v}_m^{\mathrm{T}}+I_t=0 \tag{3}$$

其中,$\nabla\boldsymbol{I}=(I_x,I_y)$ 表示点 P 的图像梯度,$\boldsymbol{v}_m=(u,v)$ 表示点 P 的光流。式(3)为光流约束方程,是基于梯度的所有光流计算方法的基础。本文利用金字塔 LK 光流法对视频序列帧进行跟踪,水平方向 X 和垂直方向 Y 的位移由跟踪结果计算。根据式 $D=\sqrt{F_y^2+F_x^2}$,计算出两帧之间的位移幅值,即动态特征,将其作为卷积神经网络的输入。

1.2　剪切波变换

将经过处理的一张二维图像输入,复合扩张仿射系统[13]可表示为:

$$SH_{\varnothing}\boldsymbol{f}(a,s,t)=\langle\boldsymbol{f},\varnothing_{a,s,t}\rangle,a>0,s\in R,\boldsymbol{t}\in\mathbb{R}^2 \tag{4}$$

其中 $\boldsymbol{f}$ 为图像,$\varnothing_{a,s,t}$ 为剪切波因子,定义为 $\varnothing_{a,s,t}(x)=|\boldsymbol{M}_{a,s}|^{-\frac{1}{2}}\varnothing(\boldsymbol{M}_{a,s}^{-1}x-\boldsymbol{t})$,$\boldsymbol{M}_{a,s}=\boldsymbol{B}_s\boldsymbol{A}_a=\begin{pmatrix}a & \sqrt{a}s\\ 0 & \sqrt{a}\end{pmatrix}$,$\boldsymbol{A}_a=\begin{pmatrix}a & 0\\ 0 & \sqrt{a}\end{pmatrix}$,$\boldsymbol{B}_s=\begin{pmatrix}1 & s\\ 0 & 1\end{pmatrix}$,$\boldsymbol{A}_a$ 是各向异性扩散矩阵,$\boldsymbol{B}_s$ 是剪切矩阵。剪切波变换具有较好的各向异性,在不同的尺度、位置和方向上进行定义,剪切波能够检测方向信息,并解释多维函数的几何性质。从人脸图像的剪切波变换到不同的子带,剪切过程如图 2 所示。

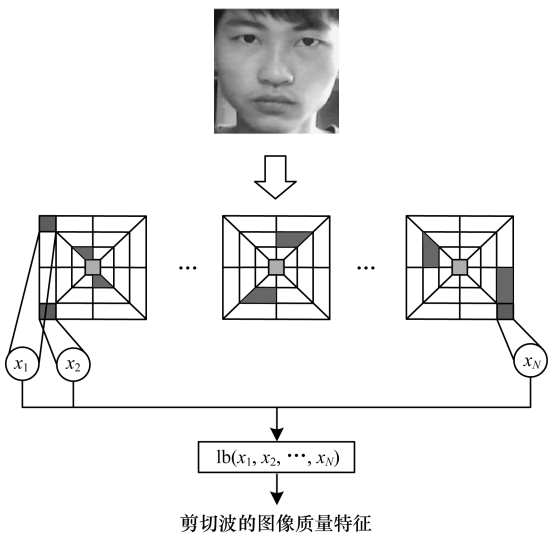

图 2　基于剪切波的图像质量特征提取过程

在图 2 中,子带中的每个元素被定义为:

$$x(a,s,b)=\frac{\sum|SH_{\varnothing}\boldsymbol{f}(a,s,b)|}{m^2} \tag{5}$$

其中,$a=1,2,\cdots,A$ 是规模指数(不包括最粗的尺度),$s=1,2,\cdots,S$ 是方向指数,$b=1,2,\cdots,(M/m)^2$ 是每个子带的块索引,M 是正方形图像的大小,m 是每个灰色方块的大小,$SH_{\varnothing}f(a,s,b)$ 为每个灰色方块的剪切波系数,并在每一个灰色方块中进行了剪切系数的平均汇集,集合的值被串联成一个对数非线性的矢量,即:

$$\boldsymbol{SBIQF}=\mathrm{lb}(x_1,x_2,\cdots,x_N) \tag{6}$$

其中，$N = A \times S \times (M/m)^2$ 是红块的总和。

1.3 卷积神经网络

传统算法需要手工设计特征，卷积神经网络则不同，需自行设计网络结构，通过训练优化网络中的权重参数即可。本文使用改进 Alex-net 卷积神经网络，结构如图 3 所示。卷积神经网络结构包括 1 个输入层、4 个卷积层、1 个全连接层和 1 个 Softmax 层[12]。其中，第 1 个卷积层和第 2 个卷积层共享权值，均包含64 个卷积核，每个卷积核的大小是 5×5，第 1 个卷积层和第 2 个卷积层的后面分别是一个最大池化层，最大池化层的大小是 3×3，第 3 个卷积层和第 4 个卷积层之间没有共享权值，卷积层分别包含 32 个卷积核，每个卷积核的大小为 3×3，全连接层由 160 个神经元组成，这些神经元与第 4 个卷积层完全相连，最后一层是 Softmax 层，包括 2 个神经元，这 2 个神经元对应于一张真实脸和一张假脸的二元分类的图像的概率分布。输入层的图像是 32 像素×32 像素，包含 RGB 的 3 个通道，对该图像进行预处理后方可进行卷积神经网络处理。一个 32 像素×32 像素的图像在 4 个角和 1 个中心的基础上被裁剪出 5 幅 24 像素×24 像素的图像，且 5 幅图像水平翻转。一幅图片被切割下来，并被翻过来得到 10 张图片。图像数量增加，即数据量增大，1 张 24 像素×24 像素的图像通过第 1 个卷积层之后得到 64 张同样大小的图像块，即特征图。在第 1 个混合池层之后，获得 64 幅 12 像素×12 像素的特征图。特征图的大小是原始数据的一半，也就是说，最大的聚合层减少了它的尺寸。特征向量的维数是原来的一半。在第 2 个卷积层和第 2 个最大池化层之后，特征图大小变为原来的一半，即 6 像素×6 像素，而第3 个与第4 个卷积层的后面没有最大池化层，特征图大小不变。在全连接层之后，一副 32 像素×32 像素的图像有 160 维特征，Softmax 层根据该特征估计概率分布。

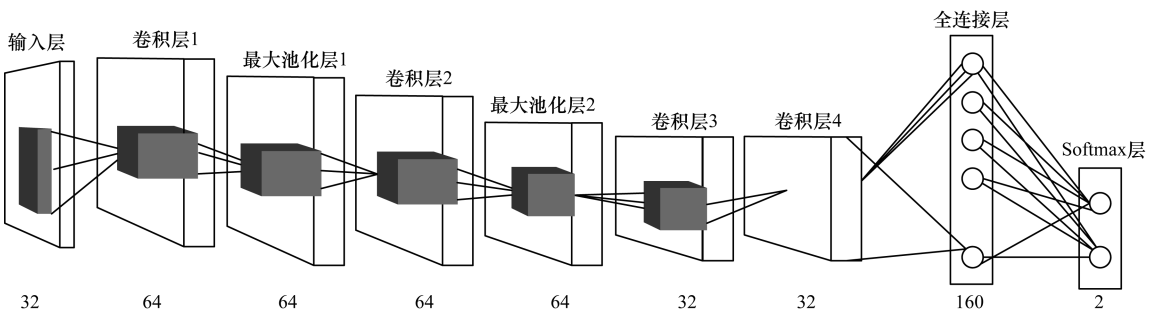

图 3 卷积神经网络结构

前向传播卷积层的计算公式为：

$$y_i = \max\{0, \sum_i w_{i,j} * x_i + b_j\} \tag{7}$$

其中，x_i 表示第 i 个输入，y_j 表示第 j 个输出，$w_{i,j}$ 表示第 i 个输入和第 j 个输出之间的卷积核，即权重参数，符号 $*$ 表示卷积运算，b_j 是 j 输出的偏移量，隐藏层使用的激活函数为整流线性单元(ReLU 函数)，其公式为 $f(x) = \max(0, x)$。最大池化层的计算公式为：

$$y_j = \max\{x_i^k\}, k \in D \tag{8}$$

其中，D 表示第 i 个输入的非重叠局部区域，y_j 是 D 中的最大值，全连接层与第 4 个卷积层完全连通，全连接层表示为：

$$y_j = \max\{0, \sum_i x_i \cdot W_{ij} + b_j\} \tag{9}$$

其中，x_i 表示第 4 个卷积层的第 i 个输入，y_j 表示全连接层的第 j 个输出，Softmax 层的输出为 n 个值，表示 n 类的概率分布。本文根据研究内容设计了 2 种输出类型，Softmax 层表示为：

$$y_j = \frac{e^{y_j'}}{\sum_j e^{y_j'}} \tag{10}$$

其中，$y_j' = \sum_{j-1}^{n} x_i \cdot w_{i,j} + b_j$，$x_i$ 表示全连接层的160 维输出，y_j 表示 Softmax 层的输出。本文只有2 个类，且 j 值是 2。

1.4 网络微调

基于卷积神经网络的人脸检测算法本质上是一个二元分类问题。而对于卷积神经网络来说，分类越多则对应的监控信号越强，鲁棒性越好，且无法从二元分类标记中直接提取相应特征。本文采用二进制分类对已训练的网络进行微调，算法框图如图 4 所示。

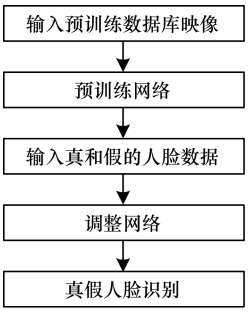

图 4 微调策略算法流程

本节图像分类包括 1 000 个类别,在预训练网络中对应 Softmax 层神经元数量为1 000。经过训练的网络会对人脸数据进行实时微调。在微调之前,Softmax 层的神经元数量应改为 2 个,对应 2 个类别,即活体和非活体。在微调后,所有网络训练结束,可以判断真假人脸。

本节所使用的卷积神经网络结构基于 Caffe 框架。Caffe 框架是一个常用的深度学习框架,是基于 C++/CUDA 架构,支持命令行、Python 和 MATLAB 接口,能够在 CPU 和 GPU 直接无缝切换,适用于人脸活体检测。

2　实验结果与分析

2.1　实验平台

本文采用 Replay-attack 数据库、CASIA-FASD 数据库,使用 Caffe 框架,编程语言使用 C++ 和 python,在基于 VS2013 开发的活体检测系统环境中对相关数据库进行测试。

2.2　实验数据库

本文采用 Replay-attack 和 CASIA-FASD 这 2 种数据库,具体描述如下:

1)Replay-attack 数据库

Replay-attack 数据库收集了 1 200 个真实拍摄的视频和 50 个用户的面部欺骗攻击视频[14]。视频在 2 种光照条件下记录:第 1 种为受控的光照条件,具有均匀的背景和人工照明;第 2 种为不利的光照条件,具有复杂的背景和自然的照明。3 种攻击包括在 A4 纸上打印的高分辨率照片、使用智能手机拍摄的视频呈现给摄像机、用平板电脑显示的高分辨率照片和视频展示给相机。2 种支持条件包括手持攻击媒体和固定攻击媒体。在 Replay-attack 数据库中,所有受试者分别被划分为 3 个非重叠子组,分别为 15 个、15 个和 20 个对象,通过训练调整分类器的参数,确定决策门槛,最后进行测试,评估分类性能。

2)CASIA-FASD 数据库

CASIA-FASD 数据库包含来自 50 个客户端的 600 个短片[15],使用 3 种不同成像的数码相机来进行拍摄以及 3 种不同的欺骗攻击。这 3 种攻击包括照片、切割图片和视频。其中,照片是指在铜质纸上印刷的脸部照片展示给摄像机,通过扭曲的照片模拟面部动作,切割图片是指眼睛区域与照片分割开来,显示眼睛闪烁,视频是指高质量的真实视频使用高分辨率的平板电脑显示。成像质量条件包括低质量、中等质量和高质量 3 种。CASIA-FASD 数据库分为训练子集和测试子集,分别有 20 个和 30 个独立的主题。

2.3　实验数据记录

定义 2 种识别失败参数,即误识率(False Rejection Rate,FRR)和拒识率(False Acceptance Rate,FAR),其计算如下:

$$FRR = \frac{\text{已拒绝但实际为活体的人脸数}}{\text{已拒绝的人脸总数}} \tag{11}$$

$$FAR = \frac{\text{已接受但实际为非活体的人脸数}}{\text{已接受的人脸数}} \tag{12}$$

本文将式(13)作为其判断算法精度的依据,与其他方法在 Replay-attack 和 CASIA-FASD 数据库上进行比较,结果如表 1、表 2 所示。

$$HTER = \frac{1}{2(FAR + FRR)} \tag{13}$$

表 1　不同方法在 Replay-attack 上性能对比　%

方法	准确率	*HTER*
DMD + LBP + SVM 方法	—	0.000
AO + Random 方法	—	0.750
lbp-top 方法	98.750	8.510
Lbp + lda 方法	—	13.870
HOOF + LDA(thresholding)方法	—	4.380
HOOF + LDA(NN) 方法	—	1.250
LFDNet 方法	97.699	3.380
CNN 方法	—	2.100
n-LBPnet 方法	—	0.017
本文方法	98.125	2.000

表 2　不同方法在 CASIA-FASD 上性能对比　%

方法	准确率	*HTER*
DMD + LBP + SVM 方法	—	0.000
AO + Random 方法	—	1.420
lbp-top 方法	97.680	9.320
Lbp + lda 方法	—	9.500
HOOF + LDA(thresholding)方法	—	3.780
HOOF + LDA(NN) 方法	—	2.420
LFDNet 方法	97.892	3.110
CNN 方法	—	2.400
n-LBPnet 方法	—	0.130
本文方法	98.891	1.750

经过上述对于不同方法在 Replay-attack 数据库和 CASIA-FASD 数据库上的准确率以及 HTER 值的比较,可见,本文方法准确率优于其他方法,且泛化力大幅提升。

3 结束语

本文提出一种基于微调策略的人脸活体检测方法。通过融合运动信息和图像质量信息,提高人脸活体检测的泛化力,利用卷积神经网络进行训练,并采用网络微调技术判断真假活体。实验结果表明,该方法相对其他方法具有较高的人脸活体检测准确度和泛化力。下一步可将动态纹理特征作为网络输入,并将其应用到并行计算上。

参考文献

[1] PAN Gang, SUN Lin, WU Zhaohui, et al. Eyeblink-based anti-spoofing in face recognition from a generic webcamera[C]//Proceedings of the 11th International Conference on Computer Vision. Washington D. C., USA: IEEE Press, 2007: 1-8.

[2] KOLLREIDER K, FRONTHALER H, FARAJ M I, et al. Real-time face detection and motion analysis with application in "liveness" assessment[J]. IEEE Transactions on Information Forensics and Security, 2007, 2(3): 548-558

[3] BAO Wei, LI Hong, LI Nan, et al. A liveness detection method for face recognition based on optical flow field[C]// Proceedings of International Conference on Image Analysis and Signal Processing. Washington D. C., USA: IEEE Press, 2009: 233-236.

[4] 任安虎,刘贝. 基于 Adaboost 的人脸识别眨眼检测[J]. 计算机与数字工程,2016,44(3):521-524.

[5] SINGH A K, JOSHI P, NANDI G C. Face recognition with liveness detection using eye and mouth movement[C]//Proceedings of International Conference on Signal Propagation and Computer Technology. Washington D. C., USA: IEEE Press, 2014: 592-597.

[6] ANJOS A, CHAKKA M M, MARCEL S. Motion-based counter-measures to photo attacks in face recognition[J]. IET Biometrics, 2014, 3(3): 147-158

[7] MAATTA J, HADID A, PIETIKAINEN M. Face spoofing detection from single images using micro-texture analysis[C]//Proceedings of International Joint Conference on Biometrics. Washington D. C., USA: IEEE Press, 2011: 1-7.

[8] 方天红,陈庆虎,廖海斌,等. 融合纹理与形状的人脸加权特征[J]. 武汉大学学报(信息科学版),2015,40(3):321-326,340.

[9] KIM Y, NA J, YOON S, et al. Masked fake face detection using radiance measurements[J]. Journal of the Optical Society of America A-Optics Image Science and Vision, 2009, 26(4): 760-766.

[10] ZHANG Zhiwei, YI Dong, LEI Zhen, et al. Face liveness detection by learning multispectral reflectance distributions[C]//Proceedings of IEEE International Conference on Automatic Face and Gesture Recognition and Workshops. Washington D. C., USA: IEEE Press, 2011: 436-441.

[11] EASLEY G, LABATE D, LIM W Q. Sparse directional image representations using the discrete shearlet transform[J]. Applied and Computational Harmonic Analysis, 2008, 25(1): 25-46.

[12] 许晓. 基于深度学习的活体人脸检测算法研究[D]. 北京:北京工业大学,2016.

[13] KUTYNIOKG, LABATE D. Shearlets: multiscale analysis for multivariate data[M]. [S.l.]: Birkhäuser Basel, 2012.

[14] CHINGOVSKA I, ANJOS A, MARCEL S. On the effectiveness of local binary patterns in face anti-spoofing[C]//Proceedings of International Conference of Biometrics Special Interest Group. Washington D. C., USA: IEEE Press, 2012: 183-194.

[15] ZHANG Zhiwei, YAN Junjie, LIU Sifei, et al. A face antispoofing database with diverse attacks[C]// Proceedings of the 5th IAPR International Conference on Biometrics. Washington D. C., USA: IEEE Press, 2012: 26-31.

编辑 赵 辉

(上接第 255 页)

[15] 李雄飞,冯婷婷,骆实,等. 基于递归神经网络的自动作曲算法[J]. 吉林大学学报(工学版),2018,48(3):866-873.

[16] HOCHREITER S, BENGIO Y, FRASCONI P, et al. Gradient flow in recurrent nets: the difficulty of learning long-term dependencies[EB/OL]. [2018-02-01]. https://www.bioinf.jku.at/publications/older/ch7.pdf.

[17] HOCHREITER S, SCHMIDHUBER J. Long short-term memory[J]. Neural Computation, 1997, 9(8): 1735-1780.

[18] LI Rongjian, ZHANG Wenlu, SUK H I, et al. Deep learning based imaging data completion for improved brain disease diagnosis[C]//Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention. Berlin, Germany: Springer, 2014: 305-312.

[19] KLERK D D. Equal temperament[J]. Acta Musicologica, 1979, 51(1): 140-150.

编辑 赵 辉