



(12) 发明专利

(10) 授权公告号 CN 111448581 B

(45) 授权公告日 2023. 12. 05

(21) 申请号 201880069485.6

(22) 申请日 2018.10.24

(65) 同一申请的已公布的文献号
申请公布号 CN 111448581 A

(43) 申请公布日 2020.07.24

(30) 优先权数据
62/576180 2017.10.24 US
62/597494 2017.12.12 US

(85) PCT国际申请进入国家阶段日
2020.04.24

(86) PCT国际申请的申请数据
PCT/CA2018/051345 2018.10.24

(87) PCT国际申请的公布数据
W02019/079895 EN 2019.05.02

(73) 专利权人 巴黎欧莱雅公司
地址 法国巴黎

(72) 发明人 艾利克斯·莱文什坦 常诚
艾德蒙·彭 埃琳娜·克泽勒
文章志·郭
艾瑞克·埃尔莫斯尼诺 蒋若为
帕汉姆·阿拉比

(74) 专利代理机构 北京泛华伟业知识产权代理
有限公司 11280
专利代理师 徐舒

(51) Int.Cl.
G06V 10/82 (2022.01)
G06V 40/16 (2022.01)
G06V 10/26 (2022.01)
G06V 20/70 (2022.01)
G06V 10/44 (2022.01)
G06T 5/20 (2006.01)
G06N 3/0464 (2023.01)
G06N 3/0455 (2023.01)
G06N 3/048 (2023.01)
G06N 3/08 (2023.01)
G06T 1/40 (2006.01)

(56) 对比文件
CN 106780512 A, 2017.05.31
CN 106203331 A, 2016.12.07
CN 105844706 A, 2016.08.10
US 2016180195 A1, 2016.06.23
JP 2016071868 A, 2016.05.09 (续)

审查员 司淼淼

权利要求书9页 说明书20页 附图10页

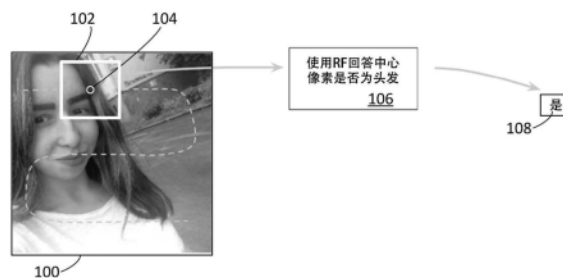
(54) 发明名称

使用深层神经网络进行图像处理的系统和方法

(57) 摘要

一种在移动设备上实现深度学习以提供用于实时处理视频,例如对头发着色,的卷积神经网络(CNN)的系统和方法。使用CNN对图像进行处理以定义头发像素的相应头发蒙版。在调整像素值时,如改变颜色、照明、纹理等时,可使用相应的物体蒙版来确定要调整的像素。该CNN可包括针对图像分类(预先训练)的适于产生分割掩膜的网络。可针对图像分割来训练该卷积神经网络(例如使用粗略分割数据训练),以最大限度地减少掩膜-图像梯度一致性损失。该CNN还可在编码

器级与解码器级的相应层之间使用跳跃连接,其中编码器中的包含高分辨率但弱的特征的较浅层与较深解码器层的低分辨率但强的特征相结合。



CN 111448581 B

[接上页]

(56) 对比文件

Siyang Qin, 等. Automatic skin and hair masking using fully convolutional networks.《2017 IEEE International Conference on Multimedia and Expo (ICME)》.2017,第103-108页.

Andrew G. Howard, 等. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications.《<https://arxiv.org/pdf/1704.04861.pdf>》.2017,第1-9页.

1. 一种处理图像的计算设备,包括:

存储单元,用于存储和提供卷积神经网络,所述卷积神经网络配置为对所述图像的像素进行分类以确定所述像素中的每个像素是否为物体像素,以便为所述图像中的物体定义物体分割掩膜,其中所述卷积神经网络包括适于定义所述物体分割掩膜的针对图像分类的预先训练的网络,其中所述卷积神经网络被进一步训练为最大限度地减少掩膜-图像梯度一致性损失,所述掩膜-图像梯度一致性损失表示在训练期间掩膜与图像梯度之间的一致性,并且其中使用分割数据进一步训练所述卷积神经网络;和

处理单元,耦合至所述存储单元,所述处理单元配置为使用所述卷积神经网络处理所述图像,以产生所述物体分割掩膜来定义新图像,

其中所述掩膜-图像梯度一致性损失 L_c 定义为:

$$L_c = \frac{\sum M_{mag} [1 - (I_x M_x + I_y M_y)^2]}{\sum M_{mag}}$$

其中 I_x 、 I_y 是归一化图像梯度, M_x 、 M_y 是归一化掩膜梯度, M_{mag} 是掩膜梯度幅值。

2. 根据权利要求1所述的计算设备,其中所述掩膜-图像梯度一致性损失 L_c 以权重 w 与二元交叉熵损失 L_M 相结合,以最大限度地减少训练时的总损失 L ,其中 L 定义为 $L = L_M + wL_c$ 。

3. 根据权利要求1至2中任一项所述的计算设备,其中所述分割数据是有噪声且粗略的分割数据。

4. 根据权利要求1至2中任一项所述的计算设备,其中所述分割数据是众包分割数据。

5. 根据权利要求1至2中任一项所述的计算设备,当在解码器级中进行上采样以定义所述物体分割掩膜时,所述计算设备适于使用编码器级中的层与解码器级中的相应层之间的跳跃连接来将低分辨率但强的特征与高分辨率但弱的特征相结合。

6. 根据权利要求1至2中任一项所述的计算设备,其中所述物体是头发。

7. 根据权利要求1至2中任一项所述的计算设备,其中所述处理单元配置为通过使用所述物体分割掩膜对选择的所述图像中的像素应用变化来从所述图像定义所述新图像。

8. 根据权利要求7所述的计算设备,其中所述卷积神经网络包括经过训练的网络,其被配置为在计算能力有限的移动设备上运行,以实时产生所述新图像作为视频图像。

9. 根据权利要求7所述的计算设备,其中所述处理单元配置为提供交互式图形用户界面以显示所述图像和所述新图像,所述交互式图形用户界面配置为接收输入以确定用于定义所述新图像的所述变化。

10. 根据权利要求9所述的计算设备,其中所述处理单元配置为使用所述物体分割掩膜来分析所述图像中的所述物体的像素,以确定用于所述变化的一个或多个候选,并且其中所述交互式图形用户界面配置为呈现所述一个或多个候选,以接收输入从而选择所述候选中的一个候选作为所述变化来应用。

11. 根据权利要求1、2、8至10中任一项所述的计算设备,其中所述图像是自拍视频。

12. 根据权利要求1、2、8至10中任一项所述的计算设备,其中所述卷积神经网络配置为使用包括深度卷积和逐点卷积的深度可分离卷积,以最大限度地减少处理操作并使得所述卷积神经网络能够在具有有限计算能力的移动设备上执行。

13. 一种处理图像的计算设备,包括:

存储单元,用于存储和提供卷积神经网络,所述卷积神经网络配置为对所述图像的像素进行分类以确定每个像素是否为头发像素,当使用分割数据训练以定义头发分割掩膜时,所述卷积神经网络被训练为最大限度地减少掩膜-图像梯度一致性损失,所述掩膜-图像梯度一致性损失表示在训练期间掩膜与图像梯度之间的一致性;和

处理单元,耦合至所述存储单元,所述处理单元配置为通过使用所述头发分割掩膜对所述图像中的头发像素应用新的头发颜色来定义和呈现着色头发图像,

其中所述掩膜-图像梯度一致性损失 L_c 定义为:

$$L_c = \frac{\sum M_{mag} [1 - (I_x M_x + I_y M_y)^2]}{\sum M_{mag}}$$

其中 I_x 、 I_y 是归一化图像梯度, M_x 、 M_y 是归一化掩膜梯度, M_{mag} 是掩膜梯度幅值。

14. 根据权利要求13所述的计算设备,其中所述卷积神经网络包括经过训练的网络,所述网络配置为在计算能力有限的移动设备上运行,以实时产生所述着色头发图像作为视频图像。

15. 根据权利要求13所述的计算设备,其中所述卷积神经网络包括适于产生所述头发分割掩膜的针对图像分类的预先训练的网络,当在解码器级中进行上采样时,所述预先训练的网络适于使用编码器级中的层与解码器级中的对应层之间的跳跃连接来将低分辨率但强的特征与高分辨率但弱的特征相结合。

16. 根据权利要求15所述的计算设备,其中通过首先应用逐点卷积来处理来自所述编码器级的较低分辨率层,以使输出深度与所述解码器级的入向的较高分辨率层相容。

17. 根据权利要求13至16中任一项所述的计算设备,其中当在解码器级中进行上采样时,所述卷积神经网络适于将来自相应层的低分辨率但强的特征与高分辨率但弱的特征相结合。

18. 根据权利要求13至16中任一项所述的计算设备,其中使用粗略分割数据训练所述卷积神经网络,以最大限度地减少掩膜-图像梯度一致性损失。

19. 根据权利要求13所述的计算设备,其中所述掩膜-图像梯度一致性损失 L_c 以权重 w 与二元交叉熵损失 L_M 相结合,以最大限度地减少训练时的总损失 L ,其中 L 定义为 $L = L_M + wL_c$ 。

20. 根据权利要求13至16、19中任一项所述的计算设备,其中所述卷积神经网络配置为使用包括深度卷积和逐点卷积的深度可分离卷积,以最大限度地减少处理操作并使得所述卷积神经网络能够在具有有限计算能力的移动设备上执行。

21. 一种处理图像的计算设备,包括:

处理单元、耦合至所述处理单元的存储单元、以及耦合至所述处理单元和所述存储单元中的至少一者的输入装置,所述存储单元存储指令,所述指令在被所述处理单元执行时配置所述计算设备,以便:

经由所述输入装置接收所述图像;

定义标识所述图像中的头发像素的头发分割掩膜;

通过使用所述头发分割掩膜对所述图像中的头发像素应用新的头发颜色来定义着色头发图像;和

提供所述着色头发图像,以输出到显示装置上;并且
其中所述头发分割掩膜是通过以下而定义的:

使用由所述存储单元存储的卷积神经网络处理所述图像,以在相继的卷积层中应用多个卷积滤波器来检测相应的特征,其中第一组相继的卷积层提供将所述图像从第一图像分辨率下采样到最小分辨率的输出,且第二组相继的卷积层将所述输出上采样回到所述第一图像分辨率,并且所述卷积神经网络被训练以最大限度地减少掩膜-图像梯度一致性损失以便定义所述头发分割掩膜,所述掩膜-图像梯度一致性损失表示在训练期间掩膜与图像梯度之间的一致性,

其中所述掩膜-图像梯度一致性损失 L_c 定义为:

$$L_c = \frac{\sum M_{mag} [1 - (I_x M_x + I_y M_y)^2]}{\sum M_{mag}}$$

其中 I_x 、 I_y 是归一化图像梯度, M_x 、 M_y 是归一化掩膜梯度, M_{mag} 是掩膜梯度幅值。

22. 根据权利要求21所述的计算设备,其中所述卷积神经网络适于使用所述第一组和所述第二组的对应卷积层之间的跳跃连接。

23. 根据权利要求21所述的计算设备,其中所述卷积神经网络包括散布在所述第二组卷积层的初始层之前和所述第二组卷积层的相应后续层之前的上采样功能,以将输出上采样到所述第一图像分辨率。

24. 根据权利要求23所述的计算设备,其中所述上采样功能使用相应的跳跃连接,每个相应的跳跃连接结合有:

第一激活图,所述第一激活图是相继的相邻卷积层的输出,作为所述第二组卷积层的下一个卷积层的输入;和

第二激活图,所述第二激活图是所述第一组卷积层中的前一个卷积层的输出,其中所述第二激活图具有比所述第一激活图大的图像分辨率。

25. 根据权利要求24所述的计算设备,其中每个相应的跳跃连接被定义为将应用于所述第二激活图的 1×1 卷积滤波器的输出与应用于所述第一激活图的上采样功能的输出相加,以将所述第一激活图的分辨率提高到较大的图像分辨率。

26. 根据权利要求21至25中任一项所述的计算设备,其中所述指令配置所述计算设备以经由所述显示装置呈现图形用户界面,所述图形用户界面包括用于观看所述图像的第一部分和用于同时观看所述着色头发图像的第二部分。

27. 根据权利要求26所述的计算设备,其中所述指令配置所述计算设备对所述着色头发图像中的所述新的头发颜色应用光照条件处理,以显示不同的光照条件下的所述新的头发颜色。

28. 根据权利要求21至25、27中任一项所述的计算设备,其中所述新的头发颜色是第一新头发颜色,并且所述着色头发图像是第一着色头发图像,并且其中所述指令配置所述计算设备以便:

通过使用所述头发分割掩膜向所述图像中的头发像素应用第二新头发颜色并通过提供第二着色头发图像以输出到所述显示装置来定义和呈现所述第二着色头发图像。

29. 根据权利要求28所述的计算设备,其中所述指令配置所述计算设备经由所述显示

装置呈现两个新颜色的图形用户界面,所述两个新颜色的图形用户界面包括用于观看所述第一着色头发图像的第一新颜色部分和用于同时观看所述第二着色头发图像的第二新颜色部分。

30. 根据权利要求26所述的计算设备,其中所述指令配置所述计算设备以便:

分析所述图像的颜色,包括所述图像中的头发像素的当前头发颜色;

确定一种或多种建议的新头发颜色;和

经由所述图形用户界面的交互部分呈现所述建议的新头发颜色,以选择所述新的头发颜色来定义所述着色头发图像。

31. 根据权利要求21至25、27、29中任一项所述的计算设备,其中所述处理单元包括用于执行所述卷积神经网络的图形处理单元GPU,并且所述计算设备包括智能手机和平板电脑之一。

32. 根据权利要求21至25、27、29中任一项所述的计算设备,其中所述图像是视频的多个视频图像之一,并且其中所述指令配置所述计算设备以处理所述多个视频图像以便改变头发颜色。

33. 根据权利要求21至25、27、29中任一项所述的计算设备,其中所述卷积神经网络包括基于MobileNet的模型,所述模型适于对所述图像的像素进行分类,以确定所述像素中的每个像素是否为头发像素。

34. 根据权利要求21至25、27、29中任一项所述的计算设备,其中所述卷积神经网络配置为包括卷积的深度可分离卷积神经网络,其中各个标准卷积被分解为深度卷积和逐点卷积,所述深度卷积限于向每个输入通道应用单个滤波器,且所述逐点卷积限于组合所述深度卷积的输出。

35. 一种配置为产生卷积神经网络的计算设备,所述卷积神经网络被训练为处理图像以定义物体分割掩膜,所述计算设备包括:

存储单元,用于接收卷积神经网络,所述卷积神经网络配置为用于图像分类并且配置为在具有有限计算能力的运行时计算设备上执行;

处理单元,配置为提供交互界面以接收输入和显示输出,所述处理单元配置为:

接收输入以调整所述卷积神经网络以便定义物体分割掩膜并向所述存储单元存储经过调整的卷积神经网络;并且

接收输入以使用被标记为用于物体分割的分割训练数据来训练经过调整的卷积神经网络,以产生所述卷积神经网络来处理图像以便定义所述物体分割掩膜,

其中所述处理单元配置为接收输入以调整所述卷积神经网络以在训练时使用为了最大限度地减少所述掩膜-图像梯度一致性损失而定义的掩膜-图像梯度一致性损失函数,所述掩膜-图像梯度一致性损失表示在训练期间掩膜与图像梯度之间的一致性,

其中所述掩膜-图像梯度一致性损失 L_c 定义为:

$$L_c = \frac{\sum M_{mag} [1 - (I_x M_x + I_y M_y)^2]}{\sum M_{mag}}$$

其中 I_x 、 I_y 是归一化图像梯度, M_x 、 M_y 是归一化掩膜梯度, M_{mag} 是掩膜梯度幅值。

36. 根据权利要求35所述的计算设备,其中所述处理单元还配置为接收输入以调整所

述卷积神经网络,以便执行以下中的至少一者:

当在解码器级中进行上采样时,使用编码器级中的层与解码器级中的相应层之间的跳跃连接来将低分辨率但强的特征与高分辨率但弱的特征相结合;和

向所述存储单元存储经过调整的卷积神经网络。

37. 一种产生被训练为处理图像以定义物体分割掩膜的卷积神经网络的方法,包括:

获得配置为用于图像分类并配置为在具有有限计算能力的计算设备上执行的卷积神经网络;

调整所述卷积神经网络以定义物体分割掩膜并且使用为了最大限度地减少所述掩膜-图像梯度一致性损失而定义的掩膜-图像梯度一致性损失函数来产生所述卷积神经网络以处理图像,所述掩膜-图像梯度一致性损失表示在训练期间掩膜与图像梯度之间的一致性;和

使用被标记为用于物体分割的分割训练数据来训练经过调整的卷积神经网络,以定义所述物体分割掩膜,

其中所述掩膜-图像梯度一致性损失 L_c 定义为:

$$L_c = \frac{\sum M_{mag} [1 - (I_x M_x + I_y M_y)^2]}{\sum M_{mag}}$$

其中 I_x 、 I_y 是归一化图像梯度, M_x 、 M_y 是归一化掩膜梯度, M_{mag} 是掩膜梯度幅值。

38. 根据权利要求37所述的方法,其中所述卷积神经网络被预先训练为用于图像分类,并且所述训练的步骤进一步训练经过预先训练的卷积神经网络。

39. 根据权利要求37至38中任一项所述的方法,其中所述分割训练数据是从众包数据定义的有噪声且粗略的分割训练数据。

40. 根据权利要求37至38中任一项所述的方法,包括调整所述卷积神经网络,以在解码器级中进行上采样时使用编码器级中的层与解码器级中的相应层之间的跳跃连接来将低分辨率但强的特征与高分辨率但弱的特征相结合。

41. 根据权利要求37至38中任一项所述的方法,其中所述掩膜-图像梯度一致性损失函数与二元交叉熵损失函数结合,以最大限度地减少综合损失。

42. 根据权利要求37至38中任一项所述的方法,包括提供卷积神经网络来处理图像,以定义存储在移动设备上的所述物体分割掩膜。

43. 根据权利要求37至38中任一项所述的方法,其中所述卷积神经网络配置为使用包括深度卷积和逐点卷积的深度可分离卷积,以最大限度地减少处理操作,并使得所述卷积神经网络能够在具有有限计算能力的移动设备上执行。

44. 一种方法,包括:

在计算设备的存储单元中存储卷积神经网络,所述卷积神经网络配置为对图像的像素进行分类,以确定所述像素中的每个像素是否为物体像素以便为所述图像中的物体定义物体分割掩膜;和

经由所述计算设备的处理单元使用所述卷积神经网络来处理图像,以产生所述物体分割掩膜,以便定义新图像;和

其中所述卷积神经网络包括适于定义所述物体分割掩膜的针对图像分类的预先训练

的网络,并且用分割数据被训练,并且

其中所述卷积神经网络被调整为在使用分割数据训练时使用掩膜-图像梯度一致性损失函数以最大限度地减少掩膜-图像梯度一致性损失,所述掩膜-图像梯度一致性损失表示在训练期间掩膜与图像梯度之间的一致性,

其中所述掩膜-图像梯度一致性损失 L_c 定义为:

$$L_c = \frac{\sum M_{mag} [1 - (I_x M_x + I_y M_y)^2]}{\sum M_{mag}}$$

其中 I_x 、 I_y 是归一化图像梯度, M_x 、 M_y 是归一化掩膜梯度, M_{mag} 是掩膜梯度幅值。

45. 根据权利要求44所述的方法,其中所述卷积神经网络还被调整为执行:

当在解码器级中进行上采样以定义所述物体分割掩膜时,使用编码器级中的层与解码器级中的相应层之间的跳跃连接来将低分辨率但强的特征与高分辨率但弱的特征相结合。

46. 根据权利要求44所述的方法,其中所述掩膜-图像梯度一致性损失以权重 w 与二元交叉熵损失相结合,以最大限度地减少训练时的总损失 L ,其中 L 定义为 $L = L_M + wL_c$ 。

47. 根据权利要求44至46中任一项所述的方法,其中所述物体是头发。

48. 根据权利要求44至46中任一项所述的方法,包括经由所述处理单元通过使用所述物体分割掩膜对选择的图像中的像素应用变化来从所述图像定义所述新图像。

49. 根据权利要求48所述的方法,其中所述变化是颜色变化。

50. 根据权利要求48所述的方法,其中所述卷积神经网络包括经过训练的网络,其被配置为在计算能力有限的移动设备上运行以实时产生所述新图像作为视频图像。

51. 根据权利要求48所述的方法,其中所述处理单元配置为提供交互式图形用户界面以显示所述图像和所述新图像,所述交互式图形用户界面配置为接收输入以确定用于定义所述新图像的所述变化。

52. 根据权利要求51所述的方法,包括经由所述处理单元使用所述物体分割掩膜来分析所述图像中的所述物体的像素,以确定用于所述变化的一个或多个候选,并且其中所述交互式图形用户界面配置为呈现所述一个或多个候选,以接收输入从而选择所述候选中的一个候选作为所述变化来应用。

53. 根据权利要求44至46、49至52中任一项所述的方法,其中所述图像是自拍视频。

54. 根据权利要求44至46、49至52中任一项所述的方法,其中所述卷积神经网络配置为使用包括深度卷积和逐点卷积的深度可分离卷积,以最大限度地减少处理操作,并使得所述卷积神经网络能够在具有有限计算能力的移动设备上执行。

55. 一种处理图像的方法,包括:

在计算设备的存储单元中存储卷积神经网络,所述卷积神经网络配置为对所述图像的像素进行分类,以确定所述像素中的每个像素是否为头发像素,当使用分割数据进行训练以定义头发分割掩膜时,所述卷积神经网络被训练为最大限度地减少掩膜-图像梯度一致性损失,所述掩膜-图像梯度一致性损失表示在训练期间掩膜与图像梯度之间的一致性;和

由所述计算设备的与所述存储单元耦合的处理单元通过使用所述头发分割掩膜向所述图像中的头发像素应用新的头发颜色来定义和呈现着色头发图像,

其中所述掩膜-图像梯度一致性损失 L_c 定义为:

$$L_c = \frac{\sum M_{mag} [1 - (I_x M_x + I_y M_y)^2]}{\sum M_{mag}}$$

其中 I_x 、 I_y 是归一化图像梯度， M_x 、 M_y 是归一化掩膜梯度， M_{mag} 是掩膜梯度幅值。

56. 根据权利要求55所述的方法，其中所述卷积神经网络包括经过训练的网络，其配置为在计算能力有限的移动设备上执行以实时产生所述着色头发图像作为视频图像。

57. 根据权利要求55所述的方法，其中所述卷积神经网络包括适于产生所述头发分割掩膜的针对图像分类的预先训练的网络，当在解码器级中进行上采样时，所述预先训练的网络适于使用编码器级中的层与解码器级中的对应层之间的跳跃连接来将低分辨率但强的特征与高分辨率但弱的特征相结合。

58. 根据权利要求57所述的方法，其中通过首先应用逐点卷积来处理来自所述编码器级的较低分辨率层，以使输出深度与所述解码器级的入向的较高分辨率层相容。

59. 根据权利要求55至58中任一项所述的方法，其中使用粗略分割数据训练所述卷积神经网络，以最大限度地减少掩膜-图像梯度一致性损失。

60. 根据权利要求55所述的方法，其中所述掩膜-图像梯度一致性损失 L_c 以权重 w 与二元交叉熵损失 L_M 相结合，以最大限度地减少训练时的总损失 L ，其中 L 定义为 $L = L_M + wL_c$ 。

61. 根据权利要求55至58中任一项所述的方法，其中所述卷积神经网络配置为使用包括深度卷积和逐点卷积的深度可分离卷积，以最大限度地减少处理操作，并使得所述卷积神经网络能够在具有有限计算能力的移动设备上执行。

62. 一种处理图像的方法，包括：

经由处理单元接收所述图像；

经由所述处理单元定义标识所述图像中的头发像素的头发分割掩膜；

经由所述处理单元通过使用所述头发分割掩膜对所述图像中的头发像素应用新的头发颜色来定义着色头发图像；和

经由所述处理单元提供所述着色头发图像以输出到显示装置上；并且

其中所述头发分割掩膜是通过以下定义的：

使用由耦合至所述处理单元的存储单元存储的卷积神经网络处理所述图像，以在相继的卷积层中应用多个卷积滤波器来检测相应的特征，并使用有噪声且粗略的分割数据来训练所述卷积神经网络，并且最大限度地减少掩膜-图像梯度一致性损失，以定义所述头发分割掩膜，所述掩膜-图像梯度一致性损失表示在训练期间掩膜与图像梯度之间的一致性，

其中所述掩膜-图像梯度一致性损失 L_c 定义为：

$$L_c = \frac{\sum M_{mag} [1 - (I_x M_x + I_y M_y)^2]}{\sum M_{mag}}$$

其中 I_x 、 I_y 是归一化图像梯度， M_x 、 M_y 是归一化掩膜梯度， M_{mag} 是掩膜梯度幅值。

63. 根据权利要求62所述方法，其中第一组相继的卷积层提供将所述图像从第一图像分辨率下采样到最小分辨率的输出，且第二组相继的卷积层将所述输出上采样回到所述第一图像分辨率，所述卷积神经网络使用所述第一组和所述第二组的相应卷积层之间的跳跃连接。

64. 根据权利要求63所述的方法，其中所述卷积神经网络包括散布在所述第二组卷积

层的初始层之前和所述第二组卷积层的相应后续层之前的上采样功能,以将输出上采样到所述第一图像分辨率。

65. 根据权利要求64所述的方法,其中所述上采样功能使用相应的跳跃连接,每个相应的跳跃连接结合有:

第一激活图,所述第一激活图是相继的相邻卷积层的输出,作为所述第二组卷积层的下一个卷积层的输入;和

第二激活图,所述第二激活图是所述第一组卷积层中的前一个卷积层的输出,其中所述第二激活图具有比所述第一激活图大的图像分辨率。

66. 根据权利要求65所述的方法,其中所述相应的跳跃连接中的每个跳跃连接被定义为将应用于所述第二激活图的 1×1 卷积滤波器的输出与应用于所述第一激活图的上采样功能的输出相加,以将所述第一激活图的分辨率提高到较大的图像分辨率。

67. 根据权利要求62至66中任一项所述的方法,包括通过所述处理单元经由所述显示装置呈现图形用户界面,所述图形用户界面包括用于观看所述图像的第一部分和用于同时观看所述着色头发图像的第二部分。

68. 根据权利要求67所述的方法,包括对所述着色头发图像中的所述新的头发颜色应用光照条件处理,以显示不同的光照条件下的所述新的头发颜色。

69. 根据权利要求62至66、68中任一项所述的方法,其中所述新的头发颜色是第一新头发颜色,并且所述着色头发图像是第一着色头发图像,并且其中所述方法包括:

由所述处理单元通过使用所述头发分割掩膜向所述图像中的头发像素应用第二新头发颜色并通过提供第二着色头发图像以输出到所述显示装置来定义和呈现所述第二着色头发图像。

70. 根据权利要求69所述的方法,包括通过所述处理单元经由所述显示装置呈现两个新颜色的图形用户界面,所述两个新颜色的图形用户界面包括用于观看所述第一着色头发图像的第一新颜色部分和用于同时观看所述第二着色头发图像的第二新颜色部分。

71. 根据权利要求67所述的方法,包括:

通过所述处理单元分析所述图像的颜色,包括所述图像中的头发像素的当前头发颜色;

确定一种或多种建议的新头发颜色;和

经由所述图形用户界面的交互部分呈现建议的新头发颜色,以选择所述新的头发颜色来定义所述着色头发图像。

72. 根据权利要求62至66、68、70、71中任一项所述的方法,其中所述处理单元包括用于执行所述卷积神经网络的图形处理单元GPU,并且所述处理单元和存储单元是由计算设备提供的,所述计算设备包括智能手机和平板电脑之一。

73. 根据权利要求62至66、68、70、71中任一项所述的方法,其中所述图像是视频的多个视频图像中的一个视频图像,并且其中所述方法包括通过所述处理单元处理所述多个视频图像以便改变头发颜色。

74. 根据权利要求62至66、68、70、71中任一项所述的方法,其中所述卷积神经网络包括基于MobileNet的模型,所述模型适于对所述图像的像素进行分类,以确定所述像素中的每个像素是否为头发像素。

75. 根据权利要求62至66、68、70、71中任一项所述的方法,其中所述卷积神经网络配置为包括卷积的深度可分离卷积神经网络,其中各个标准卷积被分解为深度卷积和逐点卷积,所述深度卷积限于向每个输入通道应用单个滤波器,而所述逐点卷积限于组合所述深度卷积的输出。

使用深层神经网络进行图像处理的系统和方法

[0001] 相互引用

[0002] 本专利申请要求以下申请在美国的国内优先权和在其他司法管辖区内的由《巴黎公约》规定的优先权:1)于2017年10月24日提交的名称为“使用深层神经网络的视频头发着色系统和方法”的美国临时申请62/576,180;以及2)于2017年12月12日提交的名称为“在移动设备上进行实时深度头发抠图的系统和方法”的美国临时申请62/597,494,在允许通过完整引用来结合的任何司法管辖区中,每个申请的全部内容通过引用结合在此。

技术领域

[0003] 本申请涉及从源图像定义新图像的图像处理,更具体地说,涉及使用诸如卷积神经网络(CNN)等深层神经网络来处理图像。

背景技术

[0004] 在许多情况下,图像处理有助于分析源图像以识别其中的特定主题,并且至少在其中一部分情况下,图像处理有助于进行校正或其他变化以产生新图像。图像处理可用于对图像中表示的物体进行分类和/或识别图像中的物体的位置。图像处理可用于校正或改变图像的属性(例如像素的相应值),例如改变颜色、纹理、照明、亮度、对比度和其他属性。其他变化可包括在图像中增加或减少物体,改变物体的形状等。

[0005] 在一个示例中,图像处理可用于对源图像中的主体的头发进行着色,以产生着色的头发图像。

[0006] 对于计算设备,尤其是常见的移动设备,例如智能手机、平板电脑等,图像处理通常要消耗大量资源。在实时处理包含一系列图像的视频时尤其如此。

发明内容

[0007] 以下说明涉及实现深度学习,尤其涉及在移动设备上实现深度学习。本公开的目的在于提供一种深度学习环境以处理实况视频,例如对头发等物体进行分割和改变头发颜色。本领域技术人员应理解,还可检测除了头发之外的物体,并且可改变颜色或其他相应属性。可使用深层神经网络处理视频图像(例如帧),以从每个视频图像定义头发像素(例如物体像素)的相应头发蒙版(例如物体掩膜)。在调整视频图像的属性(例如颜色、照明、纹理等)时,相应的物体蒙版可用于确定要调整的像素。在头发着色的一个示例中,深度学习神经网络(例如卷积神经网络)配置为对源图像的像素进行分类,以确定每个像素是否为头发像素,并定义头发掩膜。然后使用掩膜改变源图像的属性,以产生新图像。卷积神经网络可包括针对图像分类预先训练为适于产生分割掩膜的网络。可使用粗略分割数据来进一步训练卷积神经网络,并在训练时最大限度地减少掩膜-图像梯度一致性损失。卷积神经网络还可在编码器级和解码器级的相应层之间使用跳跃连接(skip connection),其中编码器中的包含高分辨率但较弱的特征的较浅层与解码器中的较深层的低分辨率但较强的特征相结合。

[0008] 这种掩膜可用于直接或间接地区分其他物体。头发掩膜可限定人体的边界或边缘,其中头发掩膜之外的主题可以是背景等。可检测的其他物体包括皮肤等。

[0009] 本发明提供了一种处理图像的计算设备,该计算设备包括:用于存储和提供卷积神经网络(CNN)的存储单元,该卷积神经网络配置为对图像的像素进行分类,以确定每个像素是否是物体像素,以便为图像中的物体定义物体分割掩膜,其中该卷积神经网络包括针对图像分类预先训练为适于定义物体分割掩膜的网络,并且其中使用分割数据进一步训练该卷积神经网络;以及耦合至存储单元的处理单元,该处理单元配置为使用卷积神经网络处理图像,以产生物体分割掩膜来定义新图像。

[0010] 当使用分割数据进行训练时,该卷积神经网络适于最大限度地减少掩膜-图像梯度一致性损失。

[0011] 当在解码器级中进行上采样以定义物体分割掩膜时,所述卷积神经网络可适于使用编码器级中的层与解码器级中的相应层之间的跳跃连接来将低分辨率但较强的特征与高分辨率但较弱的特征相结合。

[0012] 所述掩膜-图像梯度一致性损失 L_c 可定义为:

$$[0013] \quad L_c = \frac{\sum M_{mag} [1 - (I_x M_x + I_y M_y)^2]}{\sum M_{mag}}$$

[0014] 其中 I_x 、 I_y 是归一化图像梯度, M_x 、 M_y 是归一化掩膜梯度, M_{mag} 是掩膜梯度幅值。

[0015] 所述掩膜-图像梯度一致性损失可以以权重 w 与二元交叉熵损失相结合,以最大限度地减少训练时的总损失 L ,其中 L 定义为 $L = L_M + wL_c$ 。

[0016] 可使用有噪声且粗略的分割数据来训练所述卷积神经网络。所述分割数据可以是众包分割数据。

[0017] 所述物体可以是头发。所述处理单元可配置为通过使用物体分割掩膜对选择的图像中的像素应用变化来从图像定义新图像。该变化可以是颜色变化(例如当物体是头发时模拟头发颜色的变化)。

[0018] 所述卷积神经网络可包括经过训练的网络,该网络配置为在计算能力有限的移动设备上运行,以实时产生新图像作为视频图像。

[0019] 所述处理单元可配置为提供交互式图形用户界面(GUI),以显示所述图像和新图像。该交互式图形用户界面可配置为接收输入以确定用于定义新图像的变化。所述处理单元可配置为使用物体分割掩膜来分析图像中的物体的像素,以确定用于所述变化的一个或多个候选。所述交互式图形用户界面可配置为呈现一个或多个候选,以接收输入从而选择候选之一作为变化来应用。

[0020] 所述图像可以是自拍视频。

[0021] 本发明提供了一种处理图像的计算设备,该计算设备包括:用于存储和提供卷积神经网络(CNN)的存储单元,该卷积神经网络配置为对图像的像素进行分类,以确定每个像素是否为头发像素,当使用分割数据训练以定义头发分割掩膜时,所述卷积神经网络被训练为最大限度地减少掩膜-图像梯度一致性损失;以及耦合至存储单元的处理单元,该处理单元配置为通过使用头发分割掩膜对图像中的头发像素应用新的头发颜色来定义和呈现着色头发图像。

[0022] 本发明提供了一种处理图像的计算设备,该计算设备包括:处理单元、耦合至处理单元的存储单元、以及耦合至处理单元和存储单元中的至少一个的输入装置,所述存储单元存储指令,该指令在被处理单元执行时配置所述计算设备,以便:通过输入装置接收图像;定义标识图像中的头发像素的头发分割掩膜;通过使用头发分割掩膜对图像中的头发像素应用新的头发颜色来定义着色头发图像;并且提供着色头发图像,以输出到显示装置上;并且其中所述头发分割掩膜是通过以下操作定义的:使用由存储单元存储的卷积神经网络(CNN)处理图像,以在相继的卷积层中应用多个卷积(conv)滤波器来检测相应的特征,其中第一组相继的卷积层提供将图像从第一图像分辨率下采样到最小分辨率的输出,而第二组相继的卷积层将输出上采样回到第一图像分辨率;并且训练卷积神经网络以最大限度地减少掩膜-图像梯度一致性损失,以便定义头发分割掩膜。

[0023] 本发明提供了一种配置为产生卷积神经网络的计算设备,该卷积神经网络被训练为处理图像以定义物体分割掩膜,所述计算设备包括:用于接收卷积神经网络的存储单元,该卷积神经网络配置为用于图像分类,并配置为在具有有限计算能力的运行时计算设备上执行;配置为提供交互界面以接收输入和显示输出的处理单元,该处理单元配置为:接收输入以调整所述卷积神经网络,以便定义物体分割掩膜,并向存储单元存储经过调整的卷积神经网络;并且接收输入以使用被标记为用于物体分割的分割训练数据来训练经过调整的卷积神经网络,以产生卷积神经网络来处理图像,以便定义物体分割掩膜。所述处理单元可配置为接收输入以调整所述卷积神经网络,以便执行下列操作中的至少一个:在训练时使用为了最大限度地减少掩膜-图像梯度一致性损失而定义的掩膜-图像梯度一致性损失函数;并且当在解码器级中进行上采样时,使用编码器级中的层与解码器级中的相应层之间的跳跃连接来将低分辨率但较强的特征与高分辨率但较弱的特征相结合;并且向存储单元存储经过调整的卷积神经网络。

[0024] 本发明提供了一种产生被训练为处理图像以定义物体分割掩膜的卷积神经网络的方法。所述方法包括:获得配置为用于图像分类并配置为在具有有限计算能力的计算设备上执行的卷积神经网络;调整该卷积神经网络,以定义物体分割掩膜;并且使用被标记为用于物体分割的分割训练数据来训练经过调整的卷积神经网络,以定义物体分割掩膜。可使用为了最大限度地减少掩膜-图像梯度一致性损失而定义的掩膜-图像梯度一致性损失函数来训练所述卷积神经网络,以产生卷积神经网络来处理图像,以便定义物体分割掩膜。

[0025] 所述卷积神经网络可被预先训练为用于图像分类,并且所述训练步骤进一步训练经过预先训练的卷积神经网络。所述分割训练数据可以是有噪声且粗略的分割训练数据(例如从众包数据定义的分割训练数据)。

[0026] 所述方法可包括:调整所述卷积神经网络,以在解码器级中进行上采样时使用编码器级中的层与解码器级中的相应层之间的跳跃连接来将低分辨率但较强的特征与高分辨率但较弱的特征相结合。

[0027] 所述掩膜-图像梯度一致性损失函数可与二元交叉熵损失函数结合,以最大限度地减少综合损失。所述方法可包括提供卷积神经网络来处理图像,以定义将存储在移动设备上的物体分割掩膜。

[0028] 本发明提供了一种方法,该方法包括:在计算设备的存储单元中存储卷积神经网络(CNN),该卷积神经网络配置为对图像的像素进行分类,以确定每个像素是否为物体像

素,以便为图像中的物体定义物体分割掩膜,其中该卷积神经网络包括针对图像分类预先训练为适于定义物体分割掩膜的网络,并且还使用分割数据对该网络进行训练。

[0029] 当使用分割数据进行训练时,所述卷积神经网络被进一步训练为最大限度地减少掩膜-图像梯度一致性损失;并且计算设备的处理单元使用卷积神经网络处理图像,以产生物体分割掩膜,以便定义新图像。

[0030] 本发明提供了一种处理图像的方法。所述方法包括:在计算设备的存储单元中存储卷积神经网络(CNN),该卷积神经网络配置为对图像的像素进行分类,以确定每个像素是否为头发像素,当使用分割数据进行训练以定义头发分割掩膜时,该卷积神经网络被训练为最大限度地减少掩膜-图像梯度一致性损失;并且计算设备的与存储单元耦合的处理单元通过使用头发分割掩膜向图像中的头发像素应用新的头发颜色来定义和呈现着色头发图像。

[0031] 本发明提供了一种处理图像的方法,该方法包括:处理单元接收图像;处理单元定义标识图像中的头发像素的头发分割掩膜;处理单元通过使用头发分割掩膜对图像中的头发像素应用新的头发颜色来定义着色头发图像;并且处理单元提供着色头发图像,以输出到显示装置上。所述头发分割掩膜是通过以下操作定义的:使用由耦合至处理单元的存储单元存储的卷积神经网络(CNN)处理图像,以在相继的卷积层中应用多个卷积(conv)滤波器来检测相应的特征,并使用有噪声且粗略的分割数据来训练该卷积神经网络,并且最大限度地减少掩膜-图像梯度一致性损失,以定义头发分割掩膜。

[0032] 第一组相继的卷积层可提供将图像从第一图像分辨率下采样到最小分辨率的输出,而第二组相继的卷积层将输出上采样回到第一图像分辨率,所述卷积神经网络使用第一组和第二组的相应卷积层之间的跳跃连接。

[0033] 所述卷积神经网络可包括散布在第二组卷积层的初始层之前和第二组卷积层的相应后续层之前的上采样功能,以将输出上采样到第一图像分辨率。

[0034] 所述上采样功能可使用相应的跳跃连接,每个相应的跳跃连接结合了:第一激活图,该第一激活图是相继的相邻卷积层的输出,作为第二组卷积层的下一个卷积层的输入;以及第二激活图,该第二激活图是第一组卷积层中的前一个卷积层的输出,其中该第二激活图具有比第一激活图大的图像分辨率。每个相应的跳跃连接可被定义为将应用于第二激活图的 1×1 卷积滤波器的输出与应用于第一激活图的上采样功能的输出相加,以将第一激活图的分辨率提高到较大的图像分辨率。

[0035] 所述方法可包括:所述处理单元通过显示装置呈现图形用户界面(GUI),该图形用户界面包括用于观看所述图像的第一部分和用于同时观看所述着色头发图像的第二部分。

[0036] 所述方法可包括:对着色头发图像中的新头发颜色应用光照条件处理,以显示不同的光照条件下的新头发颜色。

[0037] 所述新头发颜色可以是第一新头发颜色,并且所述着色头发图像可以是第一着色头发图像。在这种情况下,所述方法可包括:所述处理装置通过使用头发分割掩膜向图像中的头发像素应用第二新头发颜色并通过提供第二着色头发图像以输出到显示装置来定义和呈现第二着色头发图像。所述方法可包括:所述处理装置通过显示装置呈现两个新颜色的图形用户界面,该两个新颜色的图形用户界面包括用于观看第一着色头发图像的第一新颜色部分和用于同时观看第二着色头发图像的第二新颜色部分。

[0038] 所述方法可包括：所述处理单元分析图像的颜色，包括图像中的头发像素的当前头发颜色；确定一种或多种建议的新头发颜色；并且通过图形用户界面的交互部分呈现建议的新头发颜色，以选择新头发颜色来定义着色头发图像。

[0039] 所述处理单元可包括用于执行卷积神经网络的图形处理单元 (GPU)，并且所述处理单元和存储单元可由计算设备提供，该计算设备包括智能手机和平板电脑之一。所述图像可以是视频的多个视频图像之一，并且所述方法可包括：所述处理单元处理多个视频图像以改变头发颜色。

[0040] 所述卷积神经网络可包括基于MobileNet的模型，该模型适于对图像的像素进行分类，以确定每个像素是否为头发像素。所述卷积神经网络可配置为包括卷积的深度可分离卷积神经网络，其中各个标准卷积被分解为深度卷积和逐点卷积，所述深度卷积限于向每个输入通道应用单个滤波器，而所述逐点卷积限于组合深度卷积的输出。

[0041] 这些和其他方面对于本领域普通技术人员来说将是显而易见的，包括计算机程序产品方面，其中 (非瞬态) 存储单元存储指令，该指令在被处理单元执行时配置计算设备的操作，以执行本文所述的通过计算机实现的任何方法方面。

附图说明

[0042] 图1是一个示例的应用基于随机森林的分类器对图像的像素进行分类的示意图；

[0043] 图2是基于谷歌公司的MobileNet模型的深层神经网络架构的示意图；

[0044] 图3是一个示例的深层神经网络架构的示意图，该深层神经网络架构配置为处理图像并产生图像掩膜作为输出；

[0045] 图4是一个示例的深层神经网络架构的示意图，该深层神经网络架构配置为使用跳跃连接来处理图像并产生图像掩膜作为输出；

[0046] 图5是一个示例的深层神经网络架构的示意图，该深层神经网络架构配置为使用跳跃连接，并采用掩膜-图像梯度一致性损失指标进行训练，以处理图像并产生图像掩膜作为输出；

[0047] 图6是图3至图5的神经网络模型的定性评估结果的图表；

[0048] 图7是图3和图5的神经网络模型的定性评估结果的图表；

[0049] 图8是一个示例的配置有深层神经网络以处理图像的计算设备的框图；

[0050] 图9是一些示例的图8的计算设备的操作流程圖；

[0051] 图10是按照本文的教导定义示例性卷积神经网络的步骤的流程图。

[0052] 通过参照附图说明本发明的一些实施例，能够最佳地说明本发明的概念，在附图中，相似的附图标记表示相似的特征。应理解，在本文中使用的术语“发明”旨在表示以下说明的实施例的基础发明概念，而不仅仅表示实施例本身。还应理解，该总体发明概念不限于以下说明的示例性实施例，并且应按照这种思想阅读以下说明。

[0053] 另外，“示例性”一词在本文中用于表示“用作示例、实例或例示”。在本文中被称为示例性实施例的任何构造、过程、设计、技术实施例不应被解读为相对于其他实施例是优选的或有利的。在本文中被示为示例性实例的任何实例的具体质量或适当性不是刻意的，也不应对其进行推断。

具体实施方式

[0054] 深度学习和分割

[0055] 实时图像分割是计算机视觉中的一个重要问题,有着广泛的应用。其中之一是在美颜应用中为实时色彩增强而进行的头发分割。但是,这种使用情况带来了额外的挑战。首先,与许多形状简单的物体不同的是,头发具有非常复杂的结构。为了实现逼真的色彩增强,粗略的头发分割掩膜是不够的。此时需要头发蒙版。其次,许多美颜应用程序运行在移动设备或网络浏览器上,但这种设备没有强大的计算资源。这使得实现实时性能变得更加困难。在本文中说明了一种在移动设备上以超过30帧/秒的速度精确分割头发的系统和方法。

[0056] 所述头发分割系统和方法基于卷积神经网络(CNN)。大多数现代卷积神经网络即使在强大的图形处理器上也不能实时运行,并且可能占用大量内存。本文的系统和方法的目标是在移动设备上实现实时性能。本发明的第一个贡献是展示了如何将谷歌公司最近提出的MobileNet™架构用于头发分割,这种架构既快速又紧凑,足以在移动设备上使用。有关MobileNet的详细信息可在Howard等人于2017年4月17日发表的“MobileNet:移动视觉应用的高效卷积神经网络”ArXiv:1704.04861v1[cs:CV]中找到,该文献通过引用结合在此。

[0057] 在没有详细客观的头发分割数据的情况下,使用有噪声且粗略的众包数据(其中标记在像素级上不是很精确的粗略分割数据)对该网络进行训练。但是,粗略分割结果对于头发色彩增强的目的来说是不太美观的。为了实现逼真的色彩增强效果,更精确的头发蒙版能产生更好的效果。本发明的第二个贡献是提出了一种无需精确的头发蒙版训练数据就能实时获得更精确的头发蒙版的方法。首先,本发明展示了如何修改基本网络架构以使其具有捕获精细细节的能力。然后,本发明展示了通过添加促进在感知上很有吸引力的抠图效果的辅助损失函数能够仅使用粗略的头发分割训练数据来训练网络,以产生详细的头发蒙版。我们将这种方法与使用简单的引导滤波器(具有与图像大小相关的线性运行时复杂度的保边滤波器)进行后处理进行了比较,并证明了它能产生更精确且更清晰的结果。

[0058] 在详细说明深度学习和分割之前,本申请人执行并评估了为移动设备上的视频内容开发头发着色方案的先前方法。例如,基于颜色直方图、位置和梯度因子的特征,结合随机森林模型(RF)开发了一种分类器。众所周知,该分类器连续处理像素,围绕图像滑动滤波器或核心,以确定滤波器中的中心像素是否为头发像素。在图1中示出了一个草图,其中通过在图像100上扫描滤波器102以使用随机森林分类器104连续处理像素,从而处理图像100,该随机森林分类器104输出滤波器102的中心像素106是否为头发像素的判断结果。可使用输出108来定义头发蒙版(未示出),其中该输出是头发蒙版中的相应像素的数据值。应理解,在图1的草图中,滤波器102、中心像素106和所示路径不是按比例绘制的。获得的结果在速度和准确性方面都没有保证,因此本发明人抛弃了这种方法,转向深度学习。

[0059] 没有首先选择深度学习的一个原因是,使其在移动设备上实时运行仍然是一个相当大的挑战。大多数深度学习架构甚至不能在强大的GPU上实时运行。最初的方法采用了视觉几何组的神经网络结构,即,VGG16。通过去除最后3层(例如全连接层和输出层)并通过添加几个卷积(在本文中通常缩写为“conv”)转置层来转换为语义分割网络,对在ImageNet(一种针对物体识别软件研究用途设计的大型可视化数据库(开源数据集))上预先训练的基于VGG16的分类网络进行了调整。虽然结果(输出)相当好,但是处理速度很慢,尤其是在

移动设备上(每帧需要超过一秒的时间)。因此,本发明人转而寻找较轻量的结构,这种结构应尺寸较小并且执行的操作较少,以提高处理速度等。

[0060] MobileNet架构

[0061] 谷歌公司的MobileNet架构是一个为移动设备实现的轻量级、预先训练的深度学习神经网络架构。图2是MobileNet网络200的示意图,示出了作为输入的源图像100、MobileNet的预先训练的卷积神经网络的网络层/层组202、以及作为MobileNet网络200的输出的类别标记204。由MobileNet处理的输入图像的分辨率为224x224像素,有3种颜色值(即,224x224x3)。

[0062] MobileNet架构采用深度可分离卷积(一种分解卷积形式)来最大限度地减少处理操作(即浮点运算、乘法和/或加法等)。深度可分离卷积将标准卷积分解(例如划分其功能)为深度卷积和1×1卷积(也称为“逐点卷积”),以通过减少或最大限度地减少所需的运算数量来使处理更快。深度卷积对每个输入通道应用一个滤波器。逐点卷积然后应用1×1卷积来组合深度卷积的输出,从而将滤波和组合功能/操作分为两个步骤,而不是由标准卷积执行的单个滤波和组合操作。因此,MobileNet网络200中的每个结构可包括两个卷积层(一个深度卷积层和一个逐点卷积层),以定义或描述“层组”。

[0063] 表1按照从左到右直到Softmax操作的顺序示出了17个层/层组202中的每个层/层组的激活图尺寸信息和处理操作信息。严格地说,从其第一个全卷积层和全连接层算起,MobileNet有28个卷积层。

层/层组	图尺寸	处理操作				
		3X3 卷积	深度卷积 3X3+BN+ReLU+Conv 1X1+BN+ReLU	7x7 平均池	全连接	Softmax
1	112x112x32	X				
2	112x112x64		X			
3-4	56x56x128		X			
5-6	28x28x256		X			
7-12	14x14x512		X			
13-14	7x7x1024		X			
15	1x1x1024			X		
16	1x1x1024				X	
17	1x1x1024					X

[0064] 表1

[0065] 在表1中,BN表示通过调整和缩放激活(例如从一个层/操作到下一个层/操作提供的激活图中的各个值)来归一化(向后续操作的)输入的批量归一化(batchnorm)函数。ReLU是一个整流器,表示整流后的线性单位函数(例如输入X的最大值函数 $\max(X, 0)$,该函数使得X的所有负值都被设置为0)。在深度卷积以及第一层中,下采样是使用跨步卷积处理的。通过7×7平均池获得的最终下采样使用基于对7×7阵列中的值求平均的下采样函数。Softmax或归一化指数函数将任意实数值的K维向量z“压缩”为实数值的K维向量 $\sigma(z)$,其中每项在范围(0,1)内,并且所有项加起来为1(例如缩放和归一化函数)。有用之处是,输出可用来表

示分类分布——在K个不同的可能结果(分类或类别)上的概率分布,因此经常与分类为K个类别的神经网络分类器结合使用。在图2中,通过处理操作对相应的17层进行灰度和图案编码。

[0067] 图3是示出具有层/层组302的经过调整的深度学习网络300的示意图。网络300还接收源图像100作为处理的输入。网络300(例如其层/层组302)适于输出头发掩膜304。表2按照从左到右的顺序示出了22个层/层组302中的每个层/层组的激活图尺寸信息和处理操作信息。

层/组	图尺寸	处理操作					
		3X3 卷 积	深度卷积 3X3+BN+ReLU+Conv 1X1+BN+ReLU	上 采 样	深卷积度 3x3+Conv 1X1+BN+ReLU	1x1 卷 积+Re LU	Soft max
1	112x112x32	X					
2	112x112x64		X				
3-4	56x56x128		X				
5-6	28x28x256		X				
7-12	28x28x512		X				
13-14	28x28x1024		X				
15	56x56x1024			X			
16	56x56x64				X		
17	112x112x64			X			
18	112x112x64				X		
19	224x224x64			X			
20	224x224x64				X		
21	224x224x2					X	
22	224x224x2						X

[0068] 表2

[0070] 网络300与MobileNet网络200类似,但是经过调整。由此避免了MobileNet网络200向下采样为 14×14 分辨率然后向下采样为 7×7 分辨率,并且在层5-14中的最小分辨率为 28×28 。去除了最后三层(即,两个全连接层和Softmax层,虽然也使用最后的Softmax层)。为了保留精细细节,通过将最后两个层的步长从2变为1来提高输出特征分辨率。由于在结合在MobileNet基础架构中的ImageNet上使用预先训练权重,因此具有更新分辨率的图层的核心相对于其原始分辨率按其比例因子进行了扩展。即,增大了2倍的层的核心被扩展了2倍,而增大了4倍的层的核心被扩展了4倍。这在编码器级中产生 28×28 最终最小分辨率。层/层组15和之后的层/层组可限定解码器级。

[0071] 层2-14、16、18和20结合有深度可分离卷积——一种分解标准卷积,其中深度卷积对每个输入通道应用单个滤波器,逐点卷积组合深度卷积的输出。深度可分离卷积具有减少计算量和模型大小的作用,这两个作用都有助于移动设备环境中的处理。

[0072] 解码器级以来自编码器级的上述卷积神经网络特征作为输入,并以原始的 224×224 分辨率将它们上采样为头发掩膜。上采样是在层15、17和19上执行的,与层16、18和20中

的进一步特征分析交替地进行。上采样由简化版本的反MobileNet架构执行。在每一级，操作通过复制 2×2 邻域中的每个像素将前一层向上采样2倍。然后，应用可分离深度卷积，接着使用64个滤波器进行逐点 1×1 卷积，然后执行ReLU操作，如表2所示。操作在层/层组21中结束，通过使用Softmax激活添加 1×1 卷积以及用于头发/非头发的2个输出通道来进行。

[0073] 虽然未示出，但是通过最大限度地减少预测掩膜与客观掩膜之间的二元交叉熵损失 L_M 来训练网络。二元交叉熵将在下文中结合图5进一步论述。

[0074] 因此，图3示出了用于头发分割的全卷积MobileNet架构。如此定义并训练的模型在相继的卷积层中应用多个卷积(conv)滤波器来检测相应的特征，其中第一组相继的卷积层提供将图像从第一图像分辨率下采样(并编码)到最小分辨率的输出，而第二组相继的卷积层将输出上采样回到第一图像分辨率。该模型的上采样操作散布在第二组卷积层的初始层之前和第二组卷积层的相应后续层之前，以将输出上采样到第一图像分辨率。第一组可限定编码器级，第二组可限定解码器级。

[0075] 训练深层神经网络需要大量数据。虽然有用一般语义分割的大型数据集，但是这些数据集对于头发分割来说不太常用。此外，与一些具有较简单的形状的物体(例如汽车)不同的是，头发的形状非常复杂。因此，为头发获得精确客观的分割更具挑战性。

[0076] 为了应对这种挑战，使用了在ImageNet上预先训练的网络。并使用头发分割数据对其进行进一步微调。但是，仍然需要数千个训练图像。数据是使用一个头发着色应用程序获得的众包数据，在该应用程序中，用户需要手工标记他们的头发。虽然获得这种数据比较廉价，但是产生的头发分割标记有很严重的噪声并且很粗略。可通过用足够好的头发掩膜处理人脸图像来手动地净化这种源数据。这比从头开始标记头发或修正不正确的分割要快得多。以相似的方式定义了两组内部测试数据。

[0077] 上述网络300是在苹果公司的结合有Core ML™库的Apple iPad Pro 12.9(2015)上实现的。Core ML自动地从MobileNet模型产生MobileNet类别，并且可按所述的方式对其进行调整。为了利用并行处理的优势，使用iPad的GPU及配套内存对该模型进行了处理。应说明的是，对于一些实现方案来说，为了实现期望的处理，可由GPU处理(例如执行)卷积神经网络，而在其他实现方案中，使用计算设备的CPU来处理就足够了。

[0078] 由于经过调整的深度学习网络(300)的紧凑性和Core ML库的使用，在单个图像上进行一次前向传导只需要60毫秒。该网络还是使用Tensorflow™(最初由谷歌公司开发的用于高性能数值计算的开源软件库，支持机器学习和深度学习)实现的。类似的处理速度较慢，大约为300毫秒。虽然Tensorflow具有NEON™优化(NEON技术是单指令多数据(SIMD)架构的高级扩展，适用于Arm Limited的某些处理器)，但它并未针对图形处理进行优化。众所周知，现代手机和平板电脑上的GPU的确具有相当大的能力。

[0079] 该模型已经产生了非常好的定性和定量结果，如表3所示，其中内部数据集1包括350个面部裁剪图像。它是从众包数据手动标注的，与所使用的训练数据类似。内部数据集2是108个3:4比例的面部图像(来自源输入装置)，并且是手动标记的。内部数据集1有较粗略的手动标记，而内部数据集2有较精细的手动标记，这两个数据集都没有精细的标记。

[0080]

测试数据集	精确率	召回率	F1分值	IoU	表现	精度
内部数据集1	0.89406	0.93423	0.91045	0.84302	0.77183	0.95994
内部数据集2	0.904836	0.92926	0.914681	0.84514	0.76855	0.95289

[0081] 表3

[0082] 但是,这种方法仍然不能捕获所有头发,并且只提供粗略且斑驳的掩膜,而不是精确的阿尔法蒙版。使用引导滤波对所得的掩膜进行后处理以使其更具视觉吸引力只能仅纠正较小的错误,如下文进一步所述。

[0083] 需要使用卷积神经网络来改善结果,以获得更真实的(优选是真实的)的抠图效果。在这个框架中存在两个挑战——在编码级,卷积神经网络300对图像进行相当严重的下采样,因此不能期望最终的掩膜包含非常高的分辨率细节。而且,如前文所述,训练数据和测试数据的精确性都不足以训练和评估抠图方法。

[0084] 为了解决关于下采样的第一个问题,在架构中增加了跳跃连接来重新定义上采样操作。通过增加跳跃连接,较强但低分辨率的特征与较弱但高分辨率的特征结合起来。需要说明的是,由于因增加了跳跃连接而不再需要限制下采样,因此该架构已恢复到原始编码器架构,一直到 7×7 分辨率。例如,若对图3的架构进行调整并且分辨率止于 28×28 ,则可使用较少的跳跃连接。这样,编码器中的较浅层(包含高分辨率但较弱的特征)与较深层的低分辨率但较强的特征相结合。可首先对入向编码器层应用 1×1 卷积,以使输出深度与入向解码器层(对于三个外部跳跃连接为64,对于内部跳跃连接为1024)相容,然后使用加法来合并这些层,由此组合这些层。对于每个分辨率,该分辨率下的最深编码器层用于跳跃连接。

[0085] 图4是示出具有层/层组400的经过调整的深度学习网络402的示意图。网络400还接收源图像100作为处理的输入。网络400(例如其层/组402)适于输出头发掩膜404。

[0086] 表4按照从左到右的顺序示出了26个层/层组402中的每个层/层组的激活图尺寸信息和处理操作信息。

[0087] 在网络300和网络400的模型中,多个卷积滤波器中的每一个产生激活图,并且由一个卷积层产生的激活图被输出为向相继的下一个卷积层提供输入。所述多个卷积滤波器包括第一组卷积滤波器和第二组卷积滤波器,从而:第一组卷积滤波器处理第一组层(例如网络400的层2-14)中的图像,因而包括编码器;并且第二组卷积滤波器处理第二组层(例如网络400中的层16、18、20、22和24)中的图像,因而包括解码器。头发分割掩膜是根据从相继的最终卷积层(例如从层25)输出的最终激活图定义的。

[0088] 网络300或网络400的模型可包括与第一组卷积滤波器搭配的相继的归一化函数和整流器函数,以对输出进行归一化和线性整流。该模型可包括与第二组卷积滤波器搭配的整流器功能,以对输出进行线性整流。在网络300或网络400的模型中,第一组层包括:由初始的 3×3 卷积滤波器限定的初始层;以及多个随后的深度可分离卷积层,每个卷积层依次由相应的深度 3×3 卷积滤波器、批量归一化函数和整流线性单元函数、 1×1 卷积滤波器、以及随后的批量归一化和整流线性单元函数限定。

[0089] 在网络300或网络400的模型中,第二组层包括:相继的多个初始层,每个初始层由相应的深度 3×3 卷积滤波器、 1×1 卷积滤波器和整流线性单元函数限定;以及由最终的 1×1 卷积滤波器和整流线性单元函数限定的最终层。

层/组	图尺寸	处理操作					
		3X3 卷积	深度卷积 3X3+BN+ReLU+Conv 1X1+BN+ReLU	上 采 样	深卷积度 3x3+Conv 1X1+BN+ReLU	1x1 卷积 +Re LU	Soft max
1	112x112x32	X					
2	112x112x64		X				
3-4	56x56x128		X				
5-6	28x28x256		X				
7-12	14x14x512		X				
13-14	7x7x1024		X				
[0090] 15	14x14x1024			X			
16	14x14x64				X		
17	28x28x64			X			
18	28x28x64				X		
19	56x56x64			X			
20	56x56x64				X		
21	112x112x64			X			
22	112x112x64				X		
23	224x224x64			X			
24	224x224x64				X		
25	224x224x2					X	
26	224x224x2						X

[0091] 表4

[0092] 图4示出了层15、17、19、21和23上的上采样操作。层15、17、19和21上的上采样包括执行与前一层的输出图的跳跃连接。前一层的图的分辨率等于目标分辨率。这些层的上采样功能对编码器级的先前的对应层的图执行 1×1 卷积操作以组合通道信息,并对通常作为输入提供给相应层的相邻层的激活图执行上采样操作(例如如上所述的 2×2),使得其具有与先前层的图相同的分辨率。增加了 1×1 卷积操作和上采样操作的输出。

[0093] 因此,上采样功能使用相应的跳跃连接,其中每个相应的跳跃连接组合了作为第二组卷积层的下一个卷积层的输入的相邻卷积层的第一激活图输出;以及第一组卷积层中的前一个卷积层的第二激活图输出,其中该第二激活图具有比第一激活图大的图像分辨率。每个相应的跳跃连接被定义为将应用于第二激活图的 1×1 卷积滤波器的输出与应用于第一激活图的上采样功能的输出相加,以将第一激活图的分辨率提高到较大的图像分辨率。最后,对最终激活图应用Softmax(归一化指数函数)来为头发分割图的每个像素定义0到1之间的值,从而定义头发分割图。

[0094] 此方法的定量评估结果在表5中示出:

[0095] 测试数据集	精确率	召回率	F1分值	IoU	表现	精度
内部数据集1	0.88237	0.94727	0.91024	0.84226	0.770518	0.95817

内部数据集2	0.908239	0.94196	0.92320	0.85971	0.79548	0.95699
--------	----------	---------	---------	---------	---------	---------

[0096] 表5

[0097] 此外,由于最终编码器分辨率的降低,上述架构即使包含额外的解码器层也快得多。在Apple iPad Pro 12.9(2015)上,在一幅图像上使用Core ML进行一次前向传导需要30毫秒。

[0098] 从精确性的角度来看,虽然第二个数据集的结果确实看起来比没有跳跃连接的模型的结果好,但是第一个数据集的结果是不确定的。这说明了上面的第二点——粗略分割数据的精确性有限,这不仅妨碍训练,也妨碍测试。

[0099] 有人认为,用这种数据进行定量评估在目前的表现水平上已经或多或少达到了它的能力。但是,定性地说,这种架构似乎也只是稍微好一点。一种可能的解释是,虽然跳跃连接架构现在有能力学习精细水平的细节,但是这些细节并不存在于我们当前的训练数据中,这使得最终的输出掩膜与训练数据集中的那些一样粗略。

[0100] 评估和最大限度地减少掩膜-图像梯度一致性损失

[0101] 鉴于现有的训练和测试数据,卷积神经网络仅限于使用粗略分割训练数据学习头发抠图。受Rhemann等人的工作的启发,增加了一个感性的掩膜正确性指标。为此,增加了掩膜与图像梯度之间的一致性指标。距离(损失)指标如下。

[0102] 掩膜-图像梯度一致性损失如公式1所示:

$$[0103] \quad L_c = \frac{\sum M_{mag} [1 - (I_x M_x + I_y M_y)^2]}{\sum M_{mag}} \quad (\text{公式1})$$

[0104] 其中 I_x 、 I_y 是归一化图像梯度, M_x 、 M_y 是归一化掩膜梯度, M_{mag} 是掩膜梯度幅值。当图像与掩膜梯度相符时,损失(L_c)的值很小。将这个损失以权重 w 加到原来的二元交叉熵损失上,构成总损失:

$$[0105] \quad L = L_M + wL_c \quad (\text{公式2})$$

[0106] 这两种损失的结合保持了训练掩膜的真实性与生成和图像边缘相符的掩膜之间的平衡。该掩膜-图像梯度一致性损失指标用于评估现有模型和训练新模型,其中二元交叉熵(损失)指标与公式1的新指标相结合,如公式2中所示。

[0107] 交叉熵损失或对数损失衡量分类模型的表现,其输出是0到1之间的概率值。随着预测概率偏离实际(真实)标记,交叉熵损失增大。在此示例中,该模型将各个观测结果 o 中的每个像素分类为两个类别 c 之一:头发或者不是头发,从而二元交叉熵损失 L_M 可按公式3计算:

$$[0108] \quad L_M = -(y \log(p) + (1-y) \log(1-p)) \quad (\text{公式3})$$

[0109] 其中 y 是观测结果 o 的正确分类的二元标记(0或1), p 是观测结果 o 属于类别 c 的预测概率。

[0110] 图5示出了包括具有跳跃连接的网络500的训练架构,该网络500接收图像100(I)并产生掩膜504(M)。网络500在结构上与网络400相当,但是使用所论述的损耗指标对其进行训练。

[0111] 还示出了输入502和输入506,输入502包括来自图像100的 I_x 、 I_y (归一化图像梯度),输入506包括输入至用于训练的掩膜-图像梯度一致性损失确定组件508的 M_x 、 M_y (归一

化掩膜梯度)。还示出二元交叉熵损失确定组件510。掩膜-图像梯度一致性损失 L_C 与二元交叉熵损失 L_M 结合(未示出),以定义损失 L 参数来训练网络500,如上所述。

[0112] 因此,网络500是包括针对图像分类预先训练的网络的卷积神经网络,例如使用开源图像训练数据预先训练的网络。该预先训练的网络适于定义物体分割掩膜(例如头发分割掩膜),而不是对图像本身进行分类。在训练时可进一步训练该卷积神经网络以最大限度地减少掩膜-图像梯度一致性损失,并且可使用粗略分割数据来训练该卷积神经网络。这种掩膜-图像梯度一致性损失可与二元交叉熵损失相结合。可进一步调整该卷积神经网络,以在解码器级中进行上采样以定义物体分割掩膜时使用编码器级中的层与解码器级中的相应层之间的跳跃连接来将低分辨率但较强的特征与高分辨率但较弱的特征相结合。

[0113] 最终的掩膜看起来更像蒙版,也更为细致。图6中的图表600示出了定性评估结果。还示出了待处理的图像602(图像100的一个示例)以及分别使用图3至图5的模型产生的各个掩膜(掩膜304、404和504的示例)。掩膜604是按照图3使用没有跳跃连接的模型(模型1)产生的,掩膜606是按照图4使用有跳跃连接的模型(模型2)产生的,掩膜608是按照图5使用有跳跃连接和训练后的掩膜-图像梯度一致性损失的模型(模型3)产生的。

[0114] 定量地说,按照新的一致性指标,这种方法的表现更好,但是按照其他指标(与客观数据的相似性),这种方法的表现较差。但是,如前文所述,考虑到现有的客观数据的准确性,可能不希望预测与客观数据的一致性超过某一水平。表6示出了使用相同测试数据集时所有模型的定性评估结果:

模式	测试数据集	精确率	召回率	F1 分值	IoU	表现	精度	掩膜-图像梯度一致性损失	客观的掩膜-图像梯度一致性损失
[0115] 1	内部数据集 1	0.89406	0.93423	0.91045	0.84302	0.77183	0.95994	0.2539	0.175
	内部数据集 2	0.904836	0.92926	0.914681	0.84514	0.76855	0.95289	0.275	0.171
2	内部数据集 1	0.88237	0.94727	0.91024	0.84226	0.770518	0.95817	0.2306	0.175
	内部数据集 2	0.908239	0.94196	0.92320	0.85971	0.79548	0.95699	0.2414	0.171
[0116] 3	内部数据集 1	0.93495	0.887203	0.89963	0.83444	0.748409	0.96119	0.095	0.175
	内部数据集 1	0.95587	0.87826	0.91198	0.842757	0.74259	0.95407	0.0912	0.171

[0117] 表6

[0118] 将图5的架构的模型3的蒙版输出与图3的架构的模型1的较粗略的掩膜输出进行了比较,并与增加了引导滤波器的模型1的输出进行了比较。引导滤波器是一种保边滤波器,并且相对于图像尺寸具有线性运行时复杂度。在iPad Pro上处理224x224的图像只需5毫秒。图7示出了描述图3和图5的模型的定性评估结果的图表700。其中示出了待处理的图像702(图像100的一个示例)。图像704是来自图3的模型1的掩膜,没有经过引导滤波器后处理。图像706是来自模型1的掩膜,其中增加了引导滤波器后处理。图像708是图5的模型3的掩膜(或蒙版)输出。如上所述,模型3使用跳跃连接,并利用来自众包数据的有噪声且粗略的分割数据对齐进行训练,并且利用掩膜-图像梯度一致性损失函数对其进行训练。

[0119] 使用引导滤波器的图像706示出了更多细节的捕获,各绺头发变得明显。但是,引导滤波器仅在掩膜边缘附近局部地添加细节。此外,细化的掩膜的边缘周围有一圈可见的光晕,当头发颜色与其周围环境的反差较低时,这圈光晕更加明显。这种光晕会在头发重新着色时导致渗色。图5的架构产生较清晰的边缘(如图像708中所示),并能捕获较长的发绺,而没有在引导滤波器后处理中看到的不应有的光晕效应。

[0120] 一个附加好处是,图5的架构的运行速度是图4的架构的两倍,在移动设备上处理每帧只需30毫秒,并且不需要额外的后处理抠图步骤。由于有助于捕捉高分辨率细节的跳跃连接的使用,图5的架构保持了原始的MobileNet编码器结构,其最深层具有7×7分辨率。这些层具有许多深度通道(1024),并且随着分辨率的提高而使处理费用变得非常高昂。与图4的架构中的28x28分辨率相比,采用7x7分辨率使处理速度更快。

[0121] 实验

[0122] 在三个数据集上对该方法进行了评估。第一个是众包数据集,包括9000个训练图像、380个验证图像和282个测试图像。所有三个数据集都包括原始图像及其翻转版本。由于目标是移动设备上的头发抠图,因此数据的预处理是通过检测面部并根据典型自拍的预期比例裁剪面部周围的区域进行的。

[0123] 为了将该方法与现有方法进行比较,评估了两个公共数据集:Kae等人的LFW部分数据集以及Guo和Aarabi的头发数据集。前者由2927个250x250图像组成,包括1500个训练图像、500个验证图像和927个测试图像。像素被标记为三类:头发、皮肤和背景,这是在超像素级产生的。后者由115个高分辨率图像组成。由于它包含的图像太少,无法用于训练,因此我们在该数据集上进行评估时使用我们的众包训练数据。为了使这个数据集与我们的训练数据一致,以类似的方式对其进行预处理(使用人脸检测和裁剪),并添加翻转图像。由于在少数情况下未检测到人脸,因此生成的数据集由212个图像组成。

[0124] 训练是采用Keras(F.Chollet等人,<https://github.com/keras-team/keras>,2015年)中的Adadelta方法(“Adadelta:一种自适应学习率方法”,arXiv预印本arXiv:1212.5701,2012)方法进行的,批量大小为4,学习率为1:0, $\rho=0:95$, $\epsilon=1e-7$ 。仅针对卷积层使用权重为 $2 \cdot 10^{-5}$ 的L2正则化。没有对深度卷积层和最后一个卷积层进行正则化。损失平衡权重设置为 $\omega=0:5$ (在公式3中)。

[0125] 在三类LFW数据中,只有头发类对掩膜-图像梯度一致性损失有作用。对该模型训练50个时期,并使用验证数据选择表现最佳的时期。在众包数据集上的训练需要在Nvidia GeForce GTX 1080Ti™(Nvidia、GeForce和GTX 1080Ti是Nvidia公司的商标)GPU上花费5个

小时,而在LFW部分上花费不到1个小时,因为训练集的大小要小得多。

[0126] A. 定量评估

[0127] 为了进行定量性能分析,测量了F1分值、表现、IoU和精度,并对所有测试图像进行平均。为了测量图像和头发掩膜边缘的一致性,还报出了掩膜-图像梯度一致性损失(公式1)。在人工清理众包图像(图像数据)时,仅对图像进行滤波,而不是相对于掩膜进行校正。因此,头发标注的质量仍然较差。因此,在对众包数据进行评估之前,对测试掩膜进行了手工校正,每次注释花费的时间不超过2分钟。这产生了稍微好一点的客观数据。在重新标记的数据上对该方法的三种变化形式(模型1、具有引导滤波器的模型1和模型3)进行了评估。所有三种方法在客观数据比较指标方面的表现相似,但是,模型3在梯度一致性损失类别中明显胜出,这表明其掩膜与图像边缘更相符。

[0128] 在LFW部分数据集上,使用Qin等人的表现最佳的方法给出了同等的表现,但它是在移动设备上实时实现的。仅使用精度指标进行评估,因为它是Qin等人使用的唯一指标。可以说,尤其是由于LFW部分是在超像素级标注的,因此其客观性可能不足以进行高精度分析。在Guo和Aarabi的数据集中,F1分值为0:9376,表现为0:8253。在这个经过后处理的数据集上重新运行HNN,获得了与作者报告的表现相似的表现,F1分值为0:7673,表现分值为0:4674。

[0129] B. 定性评估

[0130] 在公开的自拍图像上对该方法进行了评估,以进行定性分析。模型1产生了良好但粗略的掩膜。具有引导滤波器的模型1能产生更好的掩膜,但是在头发边界周围会有不应有的模糊。模型3获得了最精确和最清晰的结果。引导滤波器后处理和模型3的一个失败模式是它们对头发附近的头发状物体的分割不足,例如深色头发情况下的眉毛或浅色头发情况下的明亮背景。

[0131] 此外,头发内部的高光会导致模型3中的头发掩膜不均匀。

[0132] C. 网络架构实验

[0133] 使用验证数据对若干个解码器层通道进行了实验,但是观察到它对精度没有很大影响,根据大多数指标,64个通道产生最佳结果。这些实验是使用图4中的跳跃连接架构完成的,没有使用梯度一致性损失。

[0134] Howard等人观察到,在图像分辨率较高的情况下,MobileNet的表现更好。以精确头发抠图作为目标,使用我们的模型3进行了实验,将分辨率提高到 224×224 以上,这是在ImageNet上训练的最高分辨率MobileNet。使用模型3从 224×224 图像与 480×480 图像推断出的掩膜的定性比较表明, 480×480 的结果在头发边缘周围看起来更准确,能捕获较长的发络,包括在脸上(例如在鼻子上)的发络。但是,上一节中提到的问题也变得更严重,更多的头发掩膜渗入到非头发区域中,并且因头发的突出而使得掩膜的内部变得不均匀。此外,处理较大的图像的费用要高得多。

[0135] 如上所述,卷积神经网络配置为在用户的计算设备(例如移动设备)上以运行时方式执行。可配置为使得卷积神经网络的执行至少部分地在这样的设备的GPU上进行,以利用处理能力(例如这样的GPU中的并行处理能力)。在一些实现方案中,执行可在CPU上进行。应理解,使用粗略分割数据(训练数据)来定义训练网络的训练环境可不同于运行时环境。训练环境可能具有较高的处理能力和/或较多的存储空间来加速训练操作。

[0136] 图8是本公开的一个或多个方面的示例性计算设备800的框图,例如手持移动设备(例如智能手机或平板电脑)。但是,它可以是另一种计算设备,例如膝上型电脑、台式电脑、工作站等。如前文所述,研究和开发工作的目标是产生用于在具有有限资源的手持移动设备上操作的深层神经网络,以有效且高效地产生经过处理的图像(例如视频)。

[0137] 计算设备800例如包括用户设备,以获取一个或多个图像(例如视频)并处理图像以改变一个或多个属性,并呈现新图像。在一个示例中,对图像进行处理以改变图像中的头发的颜色。计算设备800包括一个或多个处理器802、一个或多个输入装置804、基于手势的输入/输出装置806、一个或多个通信单元808、以及一个或多个输出装置810。计算设备800还包括存储一个或多个模块和/或数据的一个或多个存储装置812。模块可包括深层神经网络模型814、具有用于图形用户界面(GUI 818)的组件的应用程序816、颜色预测组件820、以及图像获取组件822。数据可包括待处理的一个或多个图像(例如图像824)、从一个或多个图像产生的一个或多个掩膜(例如从图像824产生的掩膜826)、以及使用一个或多个掩膜和一个或多个图像产生的一个或多个新图像(例如新图像828)。

[0138] 应用程序816提供获取一个或多个图像(例如视频)、处理图像以改变一个或多个属性、并呈现新图像的功能。在一个示例中,对图像进行处理以改变图像中的头发的颜色。应用程序使用由神经网络模型814提供的深层神经网络来执行图像处理。网络模型可配置为图3、图4和图5中所示的任何模型。

[0139] 应用程序816可与某些属性数据(例如颜色数据830)相关联,以改变图像的一个或多个属性。改变属性与改变像素值以产生新图像相关。应理解,可使用各种颜色模型和数据格式来表示与图像相关的数据(例如用于存储、打印和/或显示图像),并且可相应地配置应用程序816。在其他示例中,属性数据可涉及改变效果,例如照明条件、纹理、形状等。应用程序816可配置有用于改变属性的一个或多个功能(未示出),例如向图像的期望位置应用效果(例如由深层神经网络识别的图像中的感兴趣的物体或其部分)。

[0140] 存储装置212可存储附加模块,例如操作系统832和其他模块(未示出),包括通信模块;图形处理模块(例如用于处理器802的GPU);地图模块;联系人模块;日历模块;照片/图库模块;照片(图像/媒体)编辑器;媒体播放器和/或流模块;社交媒体应用程序;浏览器模块;等等。存储装置在此可称为存储单元。

[0141] 通信通道838可耦合组件802、804、806、808、810、812中的每一个以及用于组件间通信的任何模块814、816和826,无论是以通信、物理和/或操作的方式耦合。在一些示例中,通信通道838可包括系统总线、网络连接、进程间通信数据结构或用于传送数据的任何其他方法。

[0142] 所述一个或多个处理器802可在计算设备800内实现功能和/或执行指令。例如,处理器802可配置为从存储装置812接收指令和/或数据,以执行图8所示的模块的功能以及其他功能(例如操作系统、应用程序等)。计算设备800可将数据/信息存储到存储装置812中。下面将进一步说明一些功能。应理解,操作可能不完全精确地在图8的模块814、816和826的范围内,从而一个模块可辅助另一个模块的功能。

[0143] 用于执行操作的计算机程序代码可用一种或多种程序设计语言的任何组合来编写,例如面向对象的程序设计语言(例如Java、Smalltalk、C++等)或者传统的过程编程语言(例如“C”程序设计语言或类似的程序设计语言)。

[0144] 计算设备800可产生用于在基于手势的输入/输出装置806的屏幕上显示的输出,或者,在一些示例中,用于通过投影仪、监视器或其他显示装置显示。应理解,基于手势的输入/输出装置806可使用多种技术来配置(例如对于输入能力:电阻式触摸屏、表面声波触摸屏、电容式触摸屏、投影电容式触摸屏、压敏屏、声脉冲识别触摸屏、或对其他存在敏感的屏幕技术;对于输出能力:液晶显示屏(LCD)、发光二极管显示屏(LED)、有机发光二极管显示屏(OLED)、点阵显示屏、电子墨水显示屏或类似的单色或彩色显示屏)。

[0145] 在本文所述的示例中,基于手势的输入/输出装置806包括能够从与触摸屏交互的用户接收触觉交互或手势作为输入的触摸屏装置。这种手势可包括轻击手势、拖动或划动手势、轻弹手势、暂停手势(例如用户触摸屏的相同位置至少达到一段阈值时间),其中用户触摸或指向基于手势的输入/输出装置806的一个或多个位置。基于手势的输入/输出装置806还可包括非敲击手势。基于手势的输入/输出装置806可向用户输出或显示信息,例如图形用户界面。基于手势的输入/输出装置806可呈现计算设备800的各种应用程序、功能和能力,例如包括查看图像、处理图像和显示新图像的图形用户界面818、消息应用程序、电话通信、联系人和日历应用程序、网络浏览应用程序、游戏应用程序、电子书应用程序、以及金融、支付和其他应用程序或功能等。

[0146] 虽然本公开主要以具有输入/输出能力的显示屏装置(例如触摸屏)的形式示出并论述了基于手势的输入/输出装置806,但是也可利用基于手势的输入/输出装置的其他例子,这些装置能够检测移动,并且其本身可不包括屏幕。在这种情况下,计算设备800包括显示屏,或者耦合至显示装置以呈现新图像。计算设备800可从跟踪板/触摸板、一个或多个摄像头、或其他对存在或手势敏感的输入装置接收基于手势的输入,其中所述存在指用户的存在方面,例如包括用户的全身或身体部位的运动。

[0147] 一个或多个通信单元808可经由一个或多个通信网络(未示出)通过在所述一个或多个网络上发送和/或接收网络信号与外部设备(未示出)通信,例如用于接收新的属性数据或应用功能,与另一个计算设备、打印装置或显示装置(均未示出)共享新的图像。所述通信单元可包括各种天线和/或网络接口卡、芯片(例如全球定位卫星(GPS)等),以进行无线和/或有线通信。

[0148] 输入装置804和输出装置810可包括一个或多个按钮、开关、定点装置、摄像头、键盘、麦克风、一个或多个传感器(例如生物特征传感器等)、扬声器、铃、一个或多个灯、触觉(振动)装置等之中的任何一种。这些装置中的一个或多个可通过通用串行总线(USB)或其他通信通道(例如838)耦合。摄像头(输入装置804)可以是朝向前方的(即,在同一侧),以允许用户在观看基于手势的输入/输出装置806时使用摄像头捕获图像以拍摄“自拍”照片。

[0149] 所述一个或多个存储装置812可采取不同的形式和/或配置,例如作为短期存储器或长期存储器。存储装置812可配置用于信息的短期存储,作为在断电时不保留存储的内容的易失性存储器。易失性存储器的例子包括随机存取存储器(RAM)、动态随机存取存储器(DRAM)、静态随机存取存储器(SRAM)等。在一些示例中,存储装置812还包括一个或多个计算机可读存储介质,例如与易失性存储器相比用于存储更大量的信息和/或长期存储这样的信息,在断电时保留信息。非易失性存储器的例子包括磁盘、光盘、软盘、闪存、或各种形式的电可编程存储器(EPROM)或电可擦除可编程存储器(EEPROM)。

[0150] 虽然未示出,但是计算设备可配置为训练环境,以训练神经网络模型814,例如使

用如图5所示的网络以及适当的训练和/或测试数据进行。

[0151] 计算设备800包括处理单元和耦合至该处理单元的存储单元。所述存储单元存储指令,该指令在被所述处理单元执行时配置所述计算设备以便:存储并提供深度学习神经网络模型(例如包括卷积神经网络),该深度学习神经网络模型配置为对图像的像素进行分类,以确定每个像素是否是感兴趣的物体(例如头发像素)的成员;并且通过改变作为感兴趣的物体的成员的像素的一个或多个属性来定义和呈现新图像(例如着色头发图像)。改变操作使用利用深度神经网络模型从图像定义的掩膜。在一个示例中,所述感兴趣的物体是头发,所述属性是颜色。因此,改变一个或多个属性是使用头发分割掩膜向图像中的头发像素应用新的头发颜色。

[0152] 所述深层神经网络适于计算设备的轻型架构,该计算设备是移动设备(例如智能手机或平板电脑),它具有的处理资源比“较大”的设备(例如膝上型电脑、台式电脑、工作站、服务器或其他同代计算设备)的处理资源少。

[0153] 所述深层神经网络模型可配置为包括卷积的深度可分离卷积神经网络,其中各个标准卷积被分解为深度卷积和逐点卷积。深度卷积限于对每个输入通道应用单个滤波器,而逐点卷积限于组合深度卷积的输出。

[0154] 所述深层神经网络模型还可配置为包括在编码器中的层与解码器中的相应层之间执行跳跃连接的操作,例如进行上采样时。

[0155] 可使用掩膜-图像梯度一致性损失指标来训练所述深层神经网络模型,由此相对于处理的图像和产生的掩膜来确定掩膜-图像梯度一致性损失,并且使用该损失指标来训练模型。掩膜-图像梯度一致性损失可按照公式1确定。

[0156] 图9是一个示例的计算设备800的操作900的流程图。操作900涉及计算设备800的用户使用应用程序(例如应用程序816)进行自拍,该自拍包括视频,该视频包括定义用户的头部(包括用户的头发)的图像的视频帧;选择新的头发颜色;并观看视频,在该视频中,用户的头发颜色是新的头发颜色。在操作902中,经由输入装置或单元(例如摄像头)接收图像,该输入装置或单元可调用图形用户界面818和图像获取组件822。提供图像以进行显示(904),例如提供给基于手势的输入/输出装置806或另一种显示装置。为了帮助在移动设备上处理头发,可对图像(数据)进行预处理(未示出)。可通过检测人脸并根据典型自拍的预期比例裁剪人脸周围的区域来预处理图像。

[0157] 在操作906中,经由基于手势的输入/输出装置806和图形用户界面818(交互式图形用户界面)接收输入,以选择新的头发颜色。图形用户界面818可配置为通过交互式界面呈现头发颜色数据以供选择。在一些示例中,应用程序816可配置为推荐颜色。虽然未示出,但是操作可包括根据接收的图像和可选的其他颜色(例如皮肤颜色)或灯光信息等来确定现有的或当前的头发颜色。可通过图形用户界面818请求表示为数据(未示出)的用户偏好。操作可进一步包括向颜色预测组件820提供这些数据。颜色预测组件820可具有响应于现有的头发颜色、皮肤颜色或其他颜色和/或灯光信息、用户偏好、趋势等之中的一种或多种来推荐适当的颜色(例如来自颜色数据830的新头发颜色的一个或多个候选)的功能。

[0158] 在操作908中,接收待处理的第二图像,以向第二图像中的由如本文所述的深层神经网络模型814识别的头发像素应用新的属性,即,新的头发颜色。随着摄像头不断捕获图像,用于定义现有的头发颜色的第一个图像不再是当前图像。

[0159] 在操作910中,使用模型814定义识别(第二)图像中的头发像素的头发分割掩膜。在操作912中,通过使用头发分割掩膜向图像中的头发像素应用新的头发颜色来定义新的图像(例如着色头发图像)。在操作914中,在由图形用户界面组件提供的图形用户界面中提供新图像,以输出到基于手势的输入/输出装置806上。随着从摄像头接收到更多的图像,进一步定义相应的掩膜,并且进一步定义并呈现具有着色头发的相应新图像。

[0160] 可提供附加或替代的图形用户界面或图形用户界面功能,以便于进行其他属性改变、对现有的头发颜色和新的头发颜色或两种新的头发颜色进行实时比较,或者保存显示新的头发颜色的静态图像或视频片段。操作可经由基于手势的输入/输出装置呈现图形用户界面,其中该图形用户界面例如包括以分屏排列方式布置的用于观看图像的第一部分和用于同时观看着色头发图像的第二部分。

[0161] 操作可对头发颜色(现有颜色或新颜色)应用光照条件处理,以显示不同光照条件下的头发。操作可配置为在相应的新图像中显示第一新头发颜色和第二新头发颜色。可定义单个掩膜并提供给两个独立的着色操作,以应用两种新的头发颜色。可顺序地或同时地提供第一和第二颜色的相应新着色图像。

[0162] 除了所论述的任何图形用户界面选项或控件之外,还可提供语音激活控件。

[0163] 可按类似的方式调整其他轻型架构,通过使用编码器与解码器的相应层之间的跳跃连接并使用掩膜-图像梯度一致性损失函数来对其进行训练,以产生头发分割掩膜。这种架构的一个例子是ShuffleNet™,这是一种专门为计算能力非常有限(例如10-150MFLOP)的移动设备设计的计算效率很高的卷积神经网络,它使用Zhang等人人和Megvii Technology Limited的逐点组卷积和通道混洗。在Zhang等人于2017年12月7日发表的“ShuffleNet:一种用于移动设备的极其有效的卷积神经网络”arXiv:1707.01083v2[cs:CV]中提供了关于ShuffleNet的详细信息,该文献通过引用结合在此。

[0164] 图10是按照本文的教导定义示例性卷积神经网络的步骤1000的流程图。在步骤1002中,获得针对图像分类预先训练并配置为在具有有限计算能力的计算设备上执行的卷积神经网络(例如MobileNet、ShuffleNet等)。

[0165] 在步骤1004中,调整卷积神经网络,以定义物体分割掩膜,例如移除全连接层,定义上采样操作等。在步骤1006中,调整卷积神经网络,以在解码器级中进行上采样时使用编码器级中的层与解码器级中的相应层之间的跳跃连接来将低分辨率但较强的特征与高分辨率但较弱的特征相结合。在步骤1008中,获得包括用于物体分割的标记数据的分割训练数据。如前文所述,该分割训练数据可能是众包的有噪声且粗略的分割数据,其中的物体分割掩膜(标记)不是很精细。

[0166] 在步骤1010中,定义掩膜-图像梯度一致性损失函数,作为在训练时用于最小化的参数。在步骤1012中,使用分割训练数据和掩膜-图像梯度一致性损失函数进一步训练经过预先训练的卷积神经网络,以最大限度地减少掩膜-图像梯度一致性损失,从而产生进一步训练的卷积神经网络。在步骤1014中,用包括用于物体分割的标记数据的分割测试数据测试进一步训练的卷积神经网络。很明显,图10中的一些步骤是可选的。例如但不限于:进行调整以使用跳跃连接可以是可选的。进行调整以最大限度地减少梯度一致性损失(以及类似的针对这种损失的训练)可以是可选的。所述训练数据可以是有噪声且粗略的分割数据(例如没有精细的标记),例如从众包数据获得的分割数据。

[0167] 可提供如此训练的卷积神经网络,以便存储和在所述的移动设备上使用。

[0168] 本领域普通技术人员应理解,除了计算设备方面之外,本发明还公开了计算机程序产品方面,其中指令存储在非瞬态存储装置(例如存储器、CD-ROM、DVD-ROM、光盘等)之中,以配置计算设备,以便执行在此存储的任何方法方面。

[0169] 实际的实施方案可包含本文说明的任何或所有特征。这些和其他方面、特征以及各种组合可表示为方法、设备、系统、用于执行功能的装置、程序产品,并且可按其他方式组合本文所述的特征。本文中说明了若干实施例。但是应理解,在不脱离本文说明的过程和技术的精神和范围的前提下,能够做出各种修改。此外,可在所述过程中提供其他步骤,或者从所述过程去除某些步骤,并且可向所述系统添加其他部件,或者从所述系统去除某些部件。因此,其他实施例也在下述权利要求所限定的范围之内。

[0170] 在本说明书的说明部分和权利要求中,用词“包括”和“包含”及其变化形式表示“包括但不限于”,并且它们不旨在(并且不)排除其他部件、整数或步骤。在本说明书中,除非上下文另有要求,否则单数涵盖复数。尤其是,在使用不定冠词时,除非上下文另有要求,否则本说明书应理解为考虑了复数以及单数。

[0171] 结合本发明的特定方面、实施例或示例说明的特征、整数、特性、化合物、化学成分或基团等应理解为可应用于任何其他方面、实施例或示例,除非与这些其他方面、实施例或示例不相容。本文公开的所有特征(包括任何所附权利要求、摘要和附图)和/或如此公开的任何方法或过程的所有步骤可按任何组合形式进行组合(除了至少一部分此类特征和/或步骤在其中互斥的组合形式)。本发明不限于任何前述示例或实施例的细节。本发明涵盖本说明书(包括任何所附权利要求、摘要和附图)中公开的任何新颖特征或这些特征的任何新颖组合,或者涵盖所公开的任何方法或过程的任何新颖步骤或这些步骤的任何新颖组合。

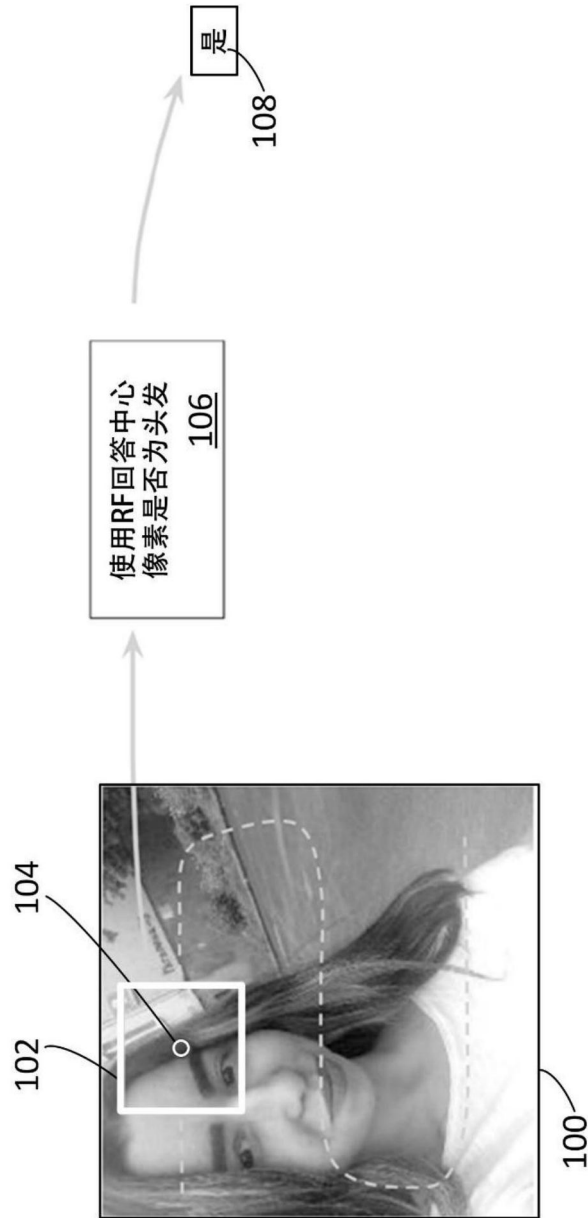


图1

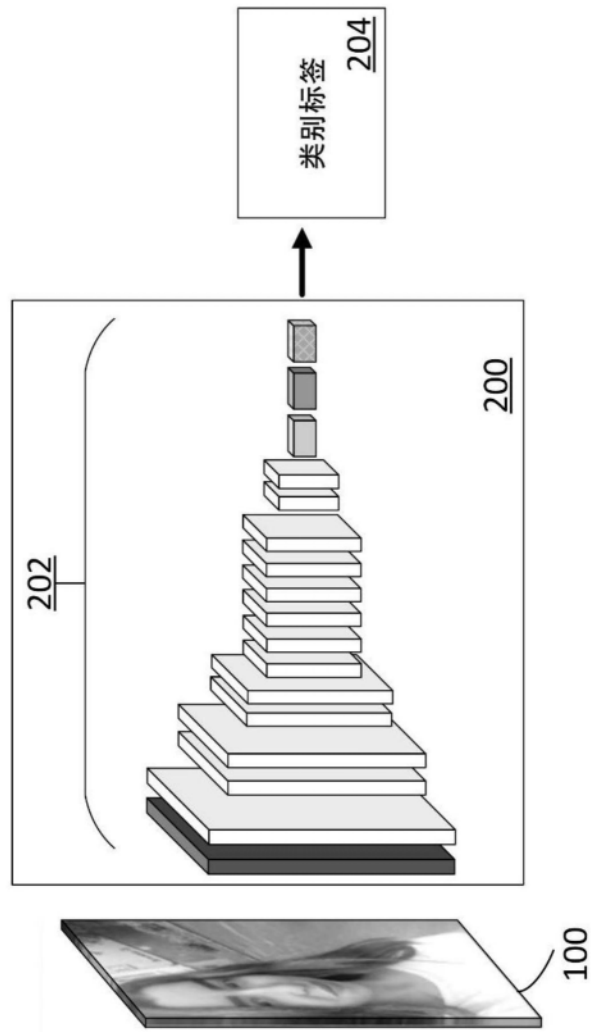


图2

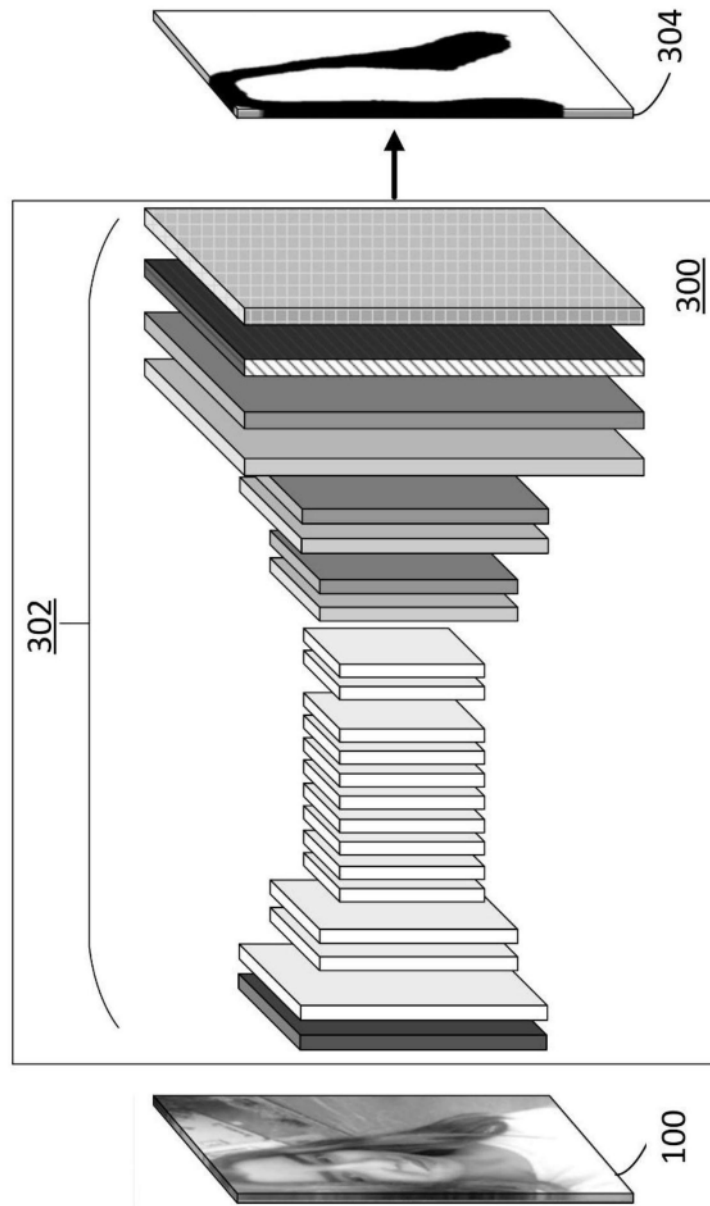


图3

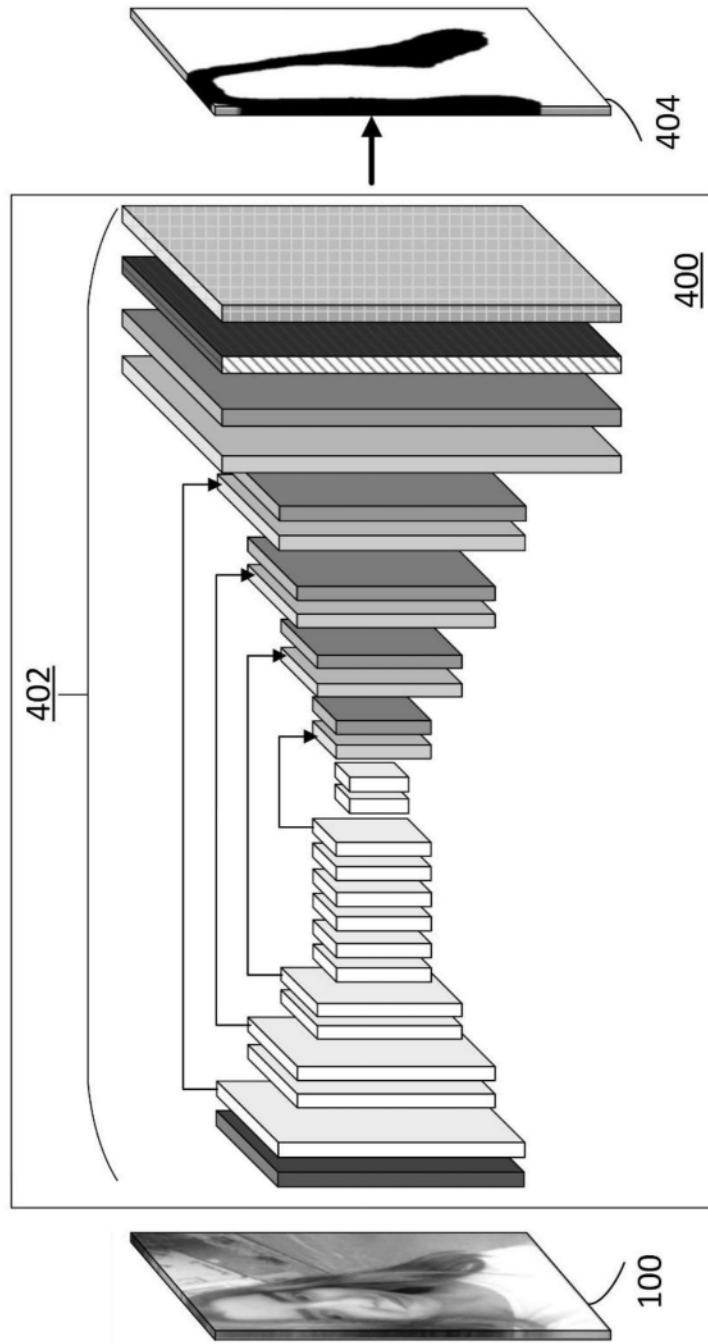


图4

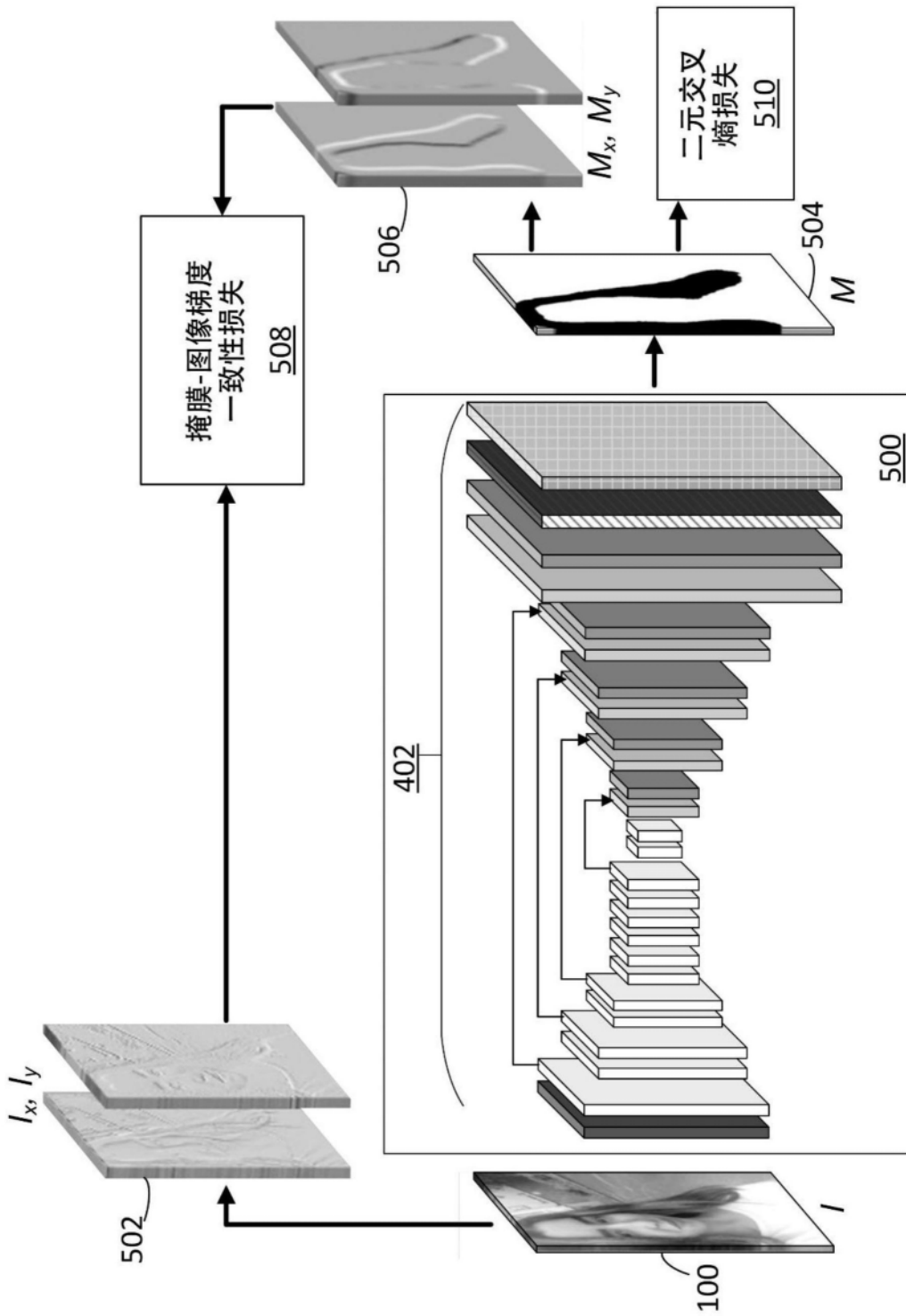


图5

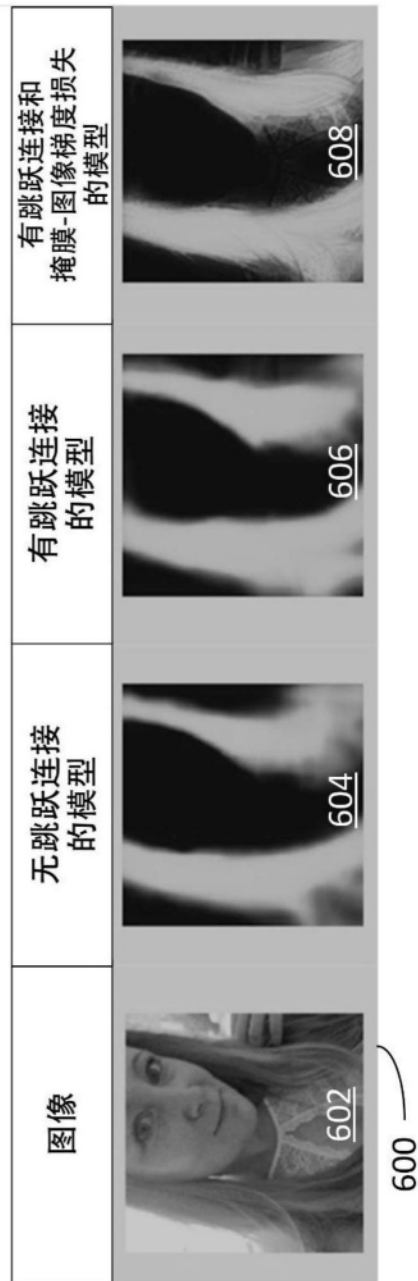


图6

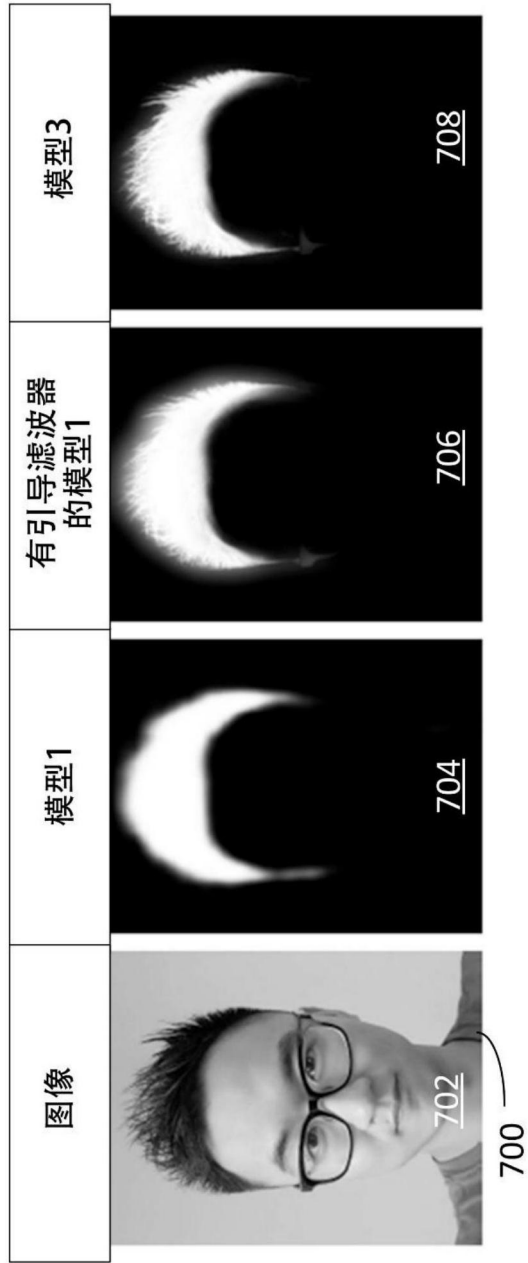


图7

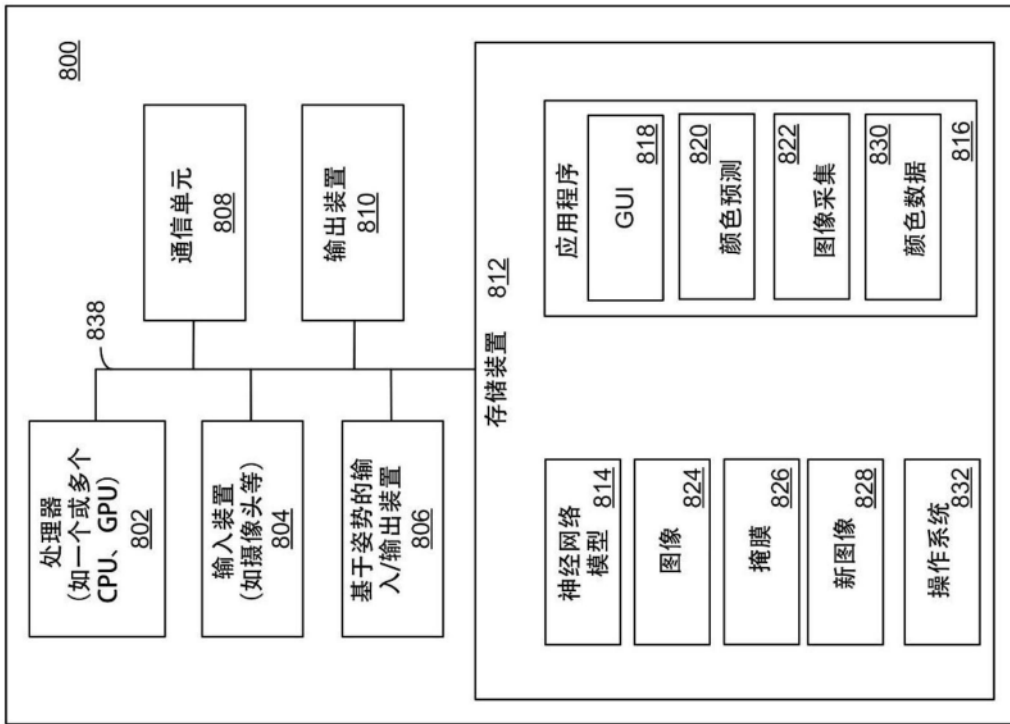


图8

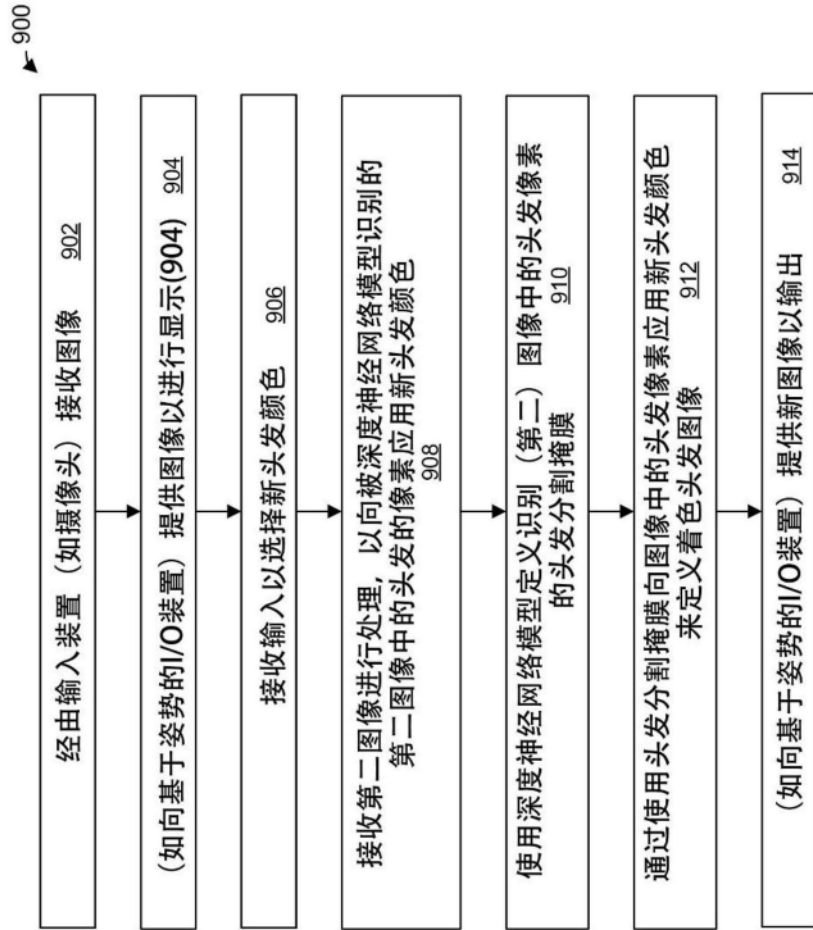


图9

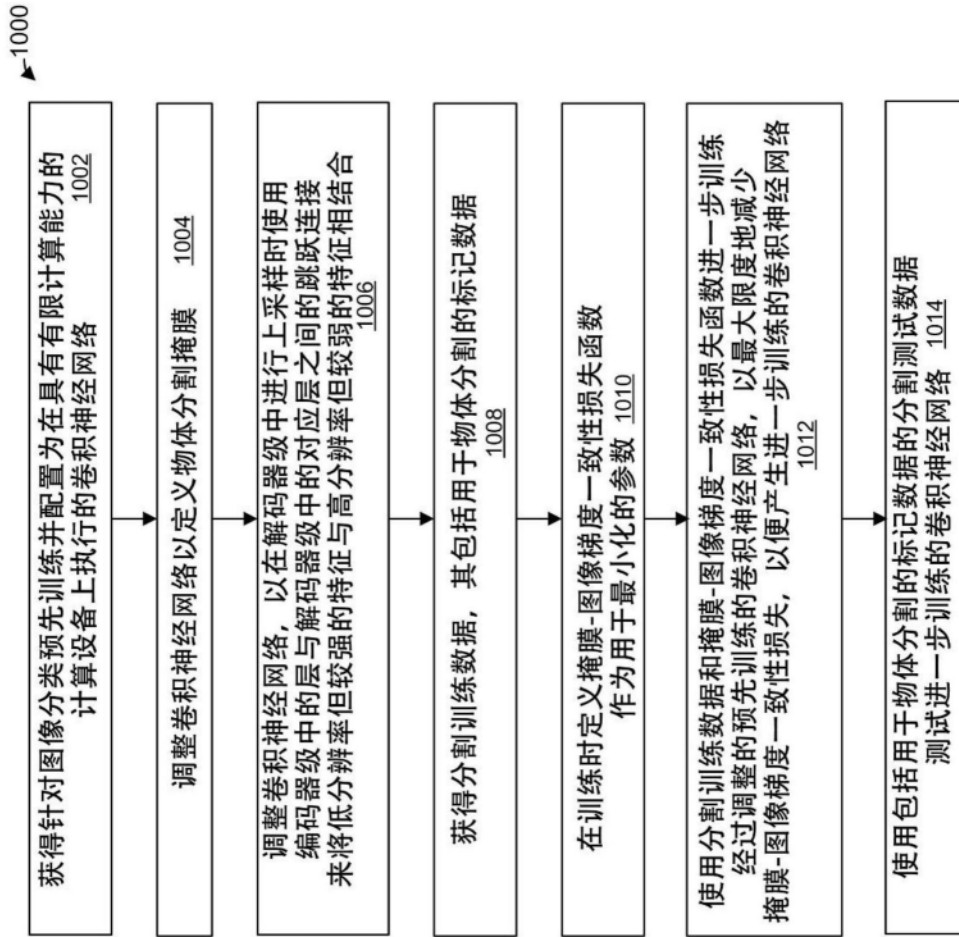


图10