

Multitask neural network design for airport surface security: A survey

nono

xxx

Abstract

首先介绍多任务的定义，必要性，以及遇到的挑战。然后，本文以视觉任务下多任务学习进行分类，主要包括场景感知的密集估计型和场景监测的非密集估计型两种。前者在多任务研究领域有较多的参考，主要涉及视觉导航类应用，而后者由于任务的抽象程度更高，工作较少，也是本文研究的重点。本文对场景监测类任务集中的多任务方法进行了总结，并提出了现有的不足以及未来可能的改进方法。最后，根据不同范式下的机器学习策略，我们对与多任务相关的联合范式学习方法进行了总结，并在现有问题上提出了可能的改进方法。

1. 引言

在机场中，塔台管制员通过目视的方法观察机场场面中航空器、车辆等目标的运行状态以及是否存在非合作目标和鸟群，获取航空器与车辆的位置、起飞降落时飞机的位姿、飞机的机尾号、机型、所属航司等信息，判断是否存在场面冲突的威胁，并进行合理的指挥控制。随着中国航空运输的快速发展，机场运输量和保障架次大幅提升，机场规模不断扩大，机场地面交通流量越来越大，人、设备等场面运行要素不断增加，导致机场运输愈加复杂，现行的目视管制方法无法感知全区域、全天候的机场运行态势。而基于视频的机场场面监视为目视管制方法的缺陷提供了有效的解决方案，实现对机场场面全方位的监控。但现有的机场场面监视方法缺少智能化，无法像管制员那样全方面获取机场中目标的运行态势，如：只能获取目标的位置。利用基于视频的监视手段辅助管制员对机场进行管理，则需要从视频中获取目标最全面的信息以及对未来运行态势的估计，如：目标的位置、速度、姿态、是否存在非合作目标以及鸟群、是否存在运行冲突，即：实现全方面的、精准的航空运行态势感知为管制员提供便利。多任务学习为智能化机场场面监视提供了有效的解决途径，通过对监控视频的分析，获取场面中目标的所有信息。而这些信息存在的内在联系也为基于机场场面安全的多任务学习提供了可能性。

人是一种能同时进行学习多种知识，也能同时执

行多种任务的生物。在进行多种知识学习的过程中，人能通过利用之前积累的知识来帮助待掌握的知识的学习，以提升效率，同时还能保证知识间的互补干扰。在执行多种任务的时候，能有效合理规划任务的安排，根据任务的不同，分配适当的时间和精力。而这种学习以及执行的方式就是未来智能化的发展方向。多任务学习（MTL）就是这样一个智能系统的雏形，Caruana在上世纪对多任务学习的进行了一个简单的描述“leveraging the domain-specific information contained in the training signals of related tasks” [1]。多任务学习认为：不同的任务之间是存在关联的，具有强耦合性的任务之间的关联性更强，而这种关联性就确定了这些任务之间的信息存在某种潜在联系，能相互补充，从而提升所有任务的泛化能力。多任务学习具有以下几个优点：1、解决训练数据稀疏的问题；2、利用已有的知识来解决未知的问题；3、通过信息、特征共享，提升特定任务的效率以及精度；4、提升所有任务的泛化能力；5、系统的算力限制导致无法多个单任务系统并行运算。可以解决由于航空运行数据采集不足导致的全面态势感知不充分以及目标定位与航空器姿态估计不准等重要问题，因此多任务学习为保障机场场面安全的未来的智能化机场场面监视方法提供了有效的解决方案。

基于机场场面安全的多任务学习仍然存在两个挑战：1）为了实现全面的航空运行态势感知，则需要实现不同任务之间的信息交互以推测未测信息；2）为了实现准确的航空运行态势感知，则需要提升所有任务的精度。这体现了全面性以及精准性，而它们严重依赖于不同任务之间的信息共享，因此，如何实现不同任务之间高效的信息交互，是基于机场场面安全的多任务学习的关键问题之一。同时，用于冲突预测的全面的、精准的数据要求所有任务的平衡，即所有任务的重要性是相同的，因此，如何实现任务之间的平衡，也是关键问题之一。基于机场场面安全的多任务学习可与其他学习方法联合起来，如迁移学习[2]、终生学习[3]、增强学习[4]。虽然多任务学习与迁移学习、终生学习等存在一定的相似之处，但是他们的目的是不同的。多任务学习强调的是多个任务整体性能的提升，迁移学习强调的是利用源域的知识提升目标域的性能，终生学习强调在新知识的学习时对已有知识的保留。并且三者在学习方式上也有同步与异步之分，即任务

的学习是否同时进行。即便如此，多任务学习、迁移学习、终生学习甚至增强学习的训练方式是可相互借鉴的。因此，利用不同的训练方式对基于机场场面安全的多任务学习进行训练也是关键问题之一。

这篇文章从四个角度对基于机场场面安全的多任务学习进行一次综述报道，首先本文对机场场面的多任务进行分类，并对场景检测类任务集进行了任务关系的层次梳理，并提出了任务关系图。其次，对目前常用的深度神经网络参数共享机制中的典型模型进行梳理，进而分析这些模型的主要优点和缺点，最后针对上述多任务深度神经网络模型存在的问题提出一些改进策略。然后，分析基于损失函数的多任务平衡方法，总结其中的难点以及一些未来可行的研究。最后，本文重点梳理其与其他学习范式的联系以及相关交叉。

2. 视觉下多任务学习中的任务关系

多任务学习中的任务，在不同的情况下意义不同，有时是代表不同尺度或者视觉层次的相同任务[5]，有时是代表一个视觉任务的不同阶段[6]，有时也代表不同种类的视觉任务[7]。本文根据应用，将多任务分成两个分支，分别是场景感知类任务集，模型多数为密集估计（Dense Prediction）的多任务集合，以及以场景监测为主，模型多数为稀疏估计的多任务集合。

2.1. 场景感知类任务集

在场景感知类任务集中，重点处理对象为场景的背景，即全局性感知，涉及任务包括深度估计[8, 9]，场景表面法向估计[10]，边缘检测[11]，光流估计[12, 13, 14]，语义分割[7, 15]，视觉里程计[16]，自运动速度估计[17]任务，图.2 和图.3 分别是两个任务以及四个任务下的多任务学习框架，可以看到每个子任务通过几何关系偶合在一起，并企图提升各自任务的效果。我们根据[18]对此类任务的层次分析，并联系涉及的应用，绘制关系如图.1

场景感知类任务集

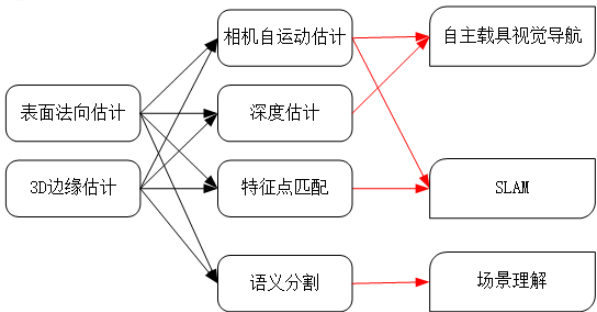


图 1. 场景感知类任务集任务关系[18]

此类任务的一个特点就是数据难以标注，人工处理困难极大等。例如，光流任务是要估计出图像中每个像素的移动，并通过色彩映射标注出来，以此估计相对相机视角下每个点的运动状况，但此种任务无法直接标注。此类任务的另一个特点就是耦合关系较强，各个任

务之间的层次性较弱。文献[18, 19]主要介绍了此类任务集的耦合关系，并量化了两两任务之间的关联度，并且还在[20]中基于前面的工作提出了如何鲁棒学习的策略，使得模型的泛化能力更好。

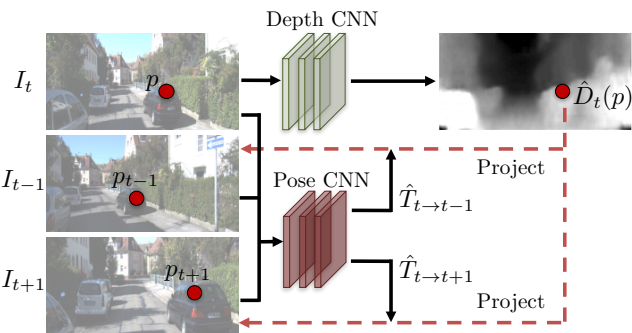


图 2. 基于深度估计任务与相机自运动位姿变换两个任务下的框图[16]

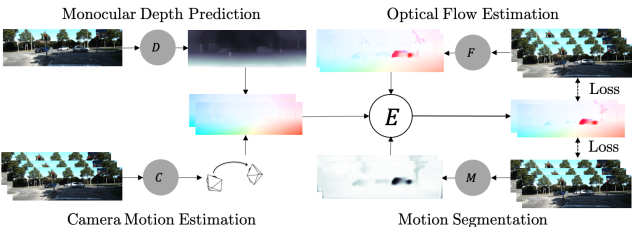


图 3. 包含了深度估计，位姿变换，分割以及光流估计四个任务的框图[7]

以上任务多数都是基于视觉导航类的任务需求，重点在探测场景感知环境上，相机一般是运动状态，算法一般搭载在机器人上，计算资源有限，与场景监测类任务有着较大的不同。

2.2. 场景监测类任务集

在场景监测类任务集中，重点处理对象是场景中的物体，即局部性感知，涉及任务包括物体检测、跟踪、运动轨迹预测、实例分割，物体位姿估计等。该任务集中的子任务抽象程度较高，没有明显的关系来关联，因此相关研究较为空白。此类任务的特点之一就是相比于场景感知类任务的数据，标注较为容易。第二个特点是此类任务的关键难点，也是本文研究的重点，即任务之间的耦合性并非通过几何耦合这类数学关系体现，而是通过逻辑关系，各个任务之间的层次性较强。因此，本文根据不同的视觉任务以及层次关系，我们将该任务集绘制如图.4

在场景检测类视觉任务衍生出的应用中，人脸检测以精度要求高，难度较大，应用前景广阔收到更大的关注。由于需要在大量人脸中准确分别，所以一些人脸检测应用中引入了更多辅助任务以增强性能。

Zhang等人[6]提出通过堆叠式多重网络来进行该任务，对人脸分类、锚箱回归、面部特征点定位作为

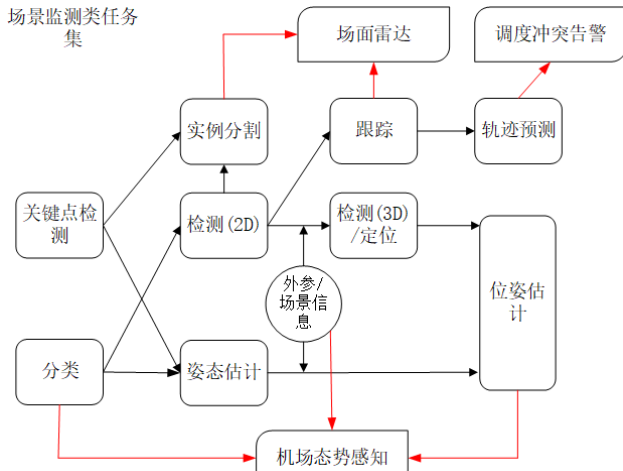


图 4. 非密集估计类任务关系

三个任务，并分别通过网络各自进行任务。但此种方法的任务实质上是一个应用级视觉任务的不同阶段，子任务的结果无法直接拿来应用，与之前提及的任务有本质不同。同样是人脸识别任务，Ranjan等[21]人以相同特征空间内提取特征并以特征提取器最后层串接多个解码器，包含了人脸检测、关键点定位、人脸姿态估计以及性别识别四个子任务，效果如图.5

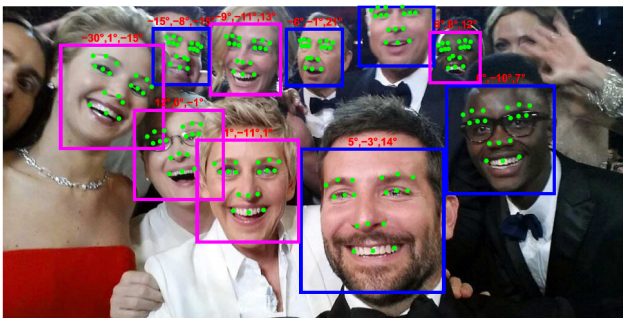


图 5. HyperFace效果展示，可见程序的结果包含人脸检测（框出）、脸部关键点定位（绿点）、性别识别（红蓝色框）以及姿态估计（俯仰偏转翻滚角度）。

综上，场景监测类任务集在的应用中，现有问题有：

- 由于缺乏耦合关系的深层次发掘，现有集中在共用特征空间的策略上，多数是一个编码器加多个解码器的结构，解码器的输入特征均来自于同一解码器的同一层，而对不同层次的特征缺少分流重建机制。
- 与场景感知类任务集基于连续可导的几何耦合关系不同，现有场景监测类任务之间暂时还没有像视图重构(View Synthesis)[22, 23]这类的成熟方法。

因此，针对问题一，根据不同视觉层次的特征来重构目标信息实现多任务学习是一个值的尝试的方法，具体来说就是后端的特征处理结构(解码器，decoder)中应该都能输入来自于多个层次的特征处理。而针对问题二，由于估计的数据结构多数为非密集型，因此结合一定的机器学习方法耦合关系挖掘应该能更好的服务于基于视觉任务的应用，又或是联系时空维度以几何方法。

3. 视觉下多任务学习外的范式关系

MTL 与机器学习中的其他学习范式相关，包括迁移学习 [24]、多标签学习 [25] 和多输出回归等。在第2章中，我们讨论了多任务学习内，任务的关系，本章则讨论多任务学习范式与其他学习范式之间的关系。

3.1. 迁移学习范式

MTL 的设置类似于迁移学习的设置，但有显著差异。在 MTL 中，不同任务之间没有区别，目标是提高所有任务的性能。然而，迁移学习是在源任务的帮助下提高目标任务的性能，因此目标任务比源任务起着更重要的作用。总之，MTL 对所有任务一视同仁，但在迁移学习中，目标任务最受关注。从知识流的角度来看，迁移学习中的知识转移流是从源任务到目标任务，但在多任务学习中，任何一对任务之间都有知识共享的流。具体来说，迁移学习是一种可以理解为持续学习的策略[26]，其中对任务的学习是依次而来，一个一个地学习任务，而MTL是一起学习多个任务。在多标签学习和多输出回归中，每个数据点都与多个标签相关联，这些标签可以是分类的或数字的。如果我们将所有可能的标签都视为一个任务，那么多标签学习和多输出回归在某种意义上可以看作是多任务学习的一个特例，其中不同的任务在训练和训练期间总是共享相同的数据。

文献[27]通过构建一个虚拟数据集，并以RCNN作为基础模型结构，载入在ImageNet上以分类为源任务的权重，在实景数据KITTI[28]上以跟踪作为目标任务，结果显示性能有着明显提升。除了源域与目标域中有限的域差异能以调优训练迁移以外，两个任务的同构性也是促进结果提升的一个重要原因。与其不同，文献[29, 30]则通过对不同数据域进行相同的多个任务初始化训练后，综合权重在一个目标域进行测试，效果在此基础上有明显的提升。为了对一些场景感知类任务达到更好的调优训练，文献[31]首次提出通过几何预训练来初始化模型而非ImageNet的分类预训练，以求在场景感知的目标任务的更好迁移，如图6所示。

3.2. 监督学习范式

在大多数应用中，标记数据的收集成本很高，但未标记的数据却很丰富。因此，在一些 MTL 应用中，每个任务的训练数据集由标记数据和未标记数据组成，因此我们希望利用未标记数据中包含的有用信息来进一步提高监督学习任务的性能。

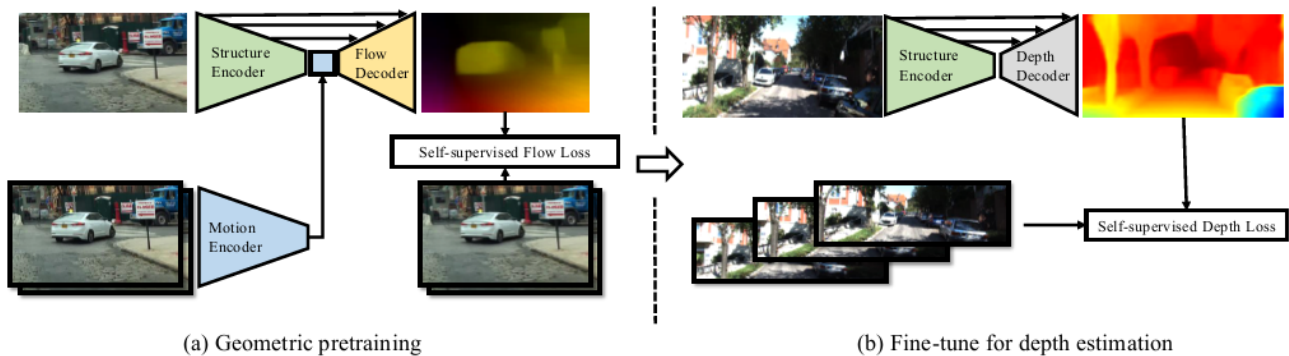


图 6. 通过场景感知类任务A进行初始化训练后, 在场景感知类任务B上进行调优, 结果相比较ImageNet有了明显的提升.

在机器学习中, 半监督学习和主动学习是利用未标记数据的两种方式, 但方式不同. 半监督学习旨在利用未标记数据中包含的几何信息, 而主动学习则选择具有代表性的未标记数据来查询预言机, 希望尽可能少地增加标记成本. 因此, 半监督学习和主动学习可以与 MTL 结合, 产生三种新的学习范式, 包括半监督多任务学习 [32, 33, 34]、多任务主动学习 [35, 36, 37] 和半监督多任务主动学习 [38]. 具体而言, [32, 33] 中提出了一种半监督多任务分类模型, 使用随机游走来利用每个任务中的未标记数据, 然后通过宽松的狄利克雷过程对多个任务进行聚类. 在 [34] 中, 提出了一种用于回归任务的半监督多任务高斯过程, 其中不同的任务通过所有任务的高斯过程中的内核参数的超先验相关联, 将未标记的数据合并到设计中每个任务中的核函数, 以在相应的功能空间中实现平滑. 与这些半监督多任务方法不同, 多任务主动学习为多任务学习者自适应地选择信息丰富的未标记数据, 因此选择标准是核心研究问题. Richard 等人. [35] 相信要选择的数据实例应该为一组任务提供尽可能多的信息, 而不是只有一个任务, 因此他们提出了两种用于多任务主动学习的协议. 在 [36] 中, 预期误差减少用作标准, 其中每个任务都由监督的潜在狄利克雷分配模型建模. 受平衡开发和探索之间权衡的多臂老虎机的启发, [37] 中提出了一种选择策略, 以考虑基于轨迹范数正则化的多任务学习器的风险和相应的置信度边界. 在 [39] 中, 提出的泛化界限用于从多个未标记的任务中选择一个子集来获取标签, 以提高所有任务的泛化性能. 对于半监督多任务主动学习, Li 等人. [38] 提出了一个模型, 使用 Fisher 信息作为标准来选择未标记的数据以获取它们的标签, 其中半监督多任务分类模型 [32, 33] 作为每个任务的分类器.

3.3. 无监督学习范式

MTL 不仅在监督学习任务中实现了性能提升, 而且一些无监督, 或者自监督的任务中也起到了一些作用. 文献 [11] 通过 3D-ASAP 规则将表面法向和深度信息联系起来, 并以深度估计为主要任务进行无监督训练, 如图 7. 文献 [14] 则利用了光流信息, 对深度估计模型进行无监督信号的加强, 主要是以前后向光流的一致性得

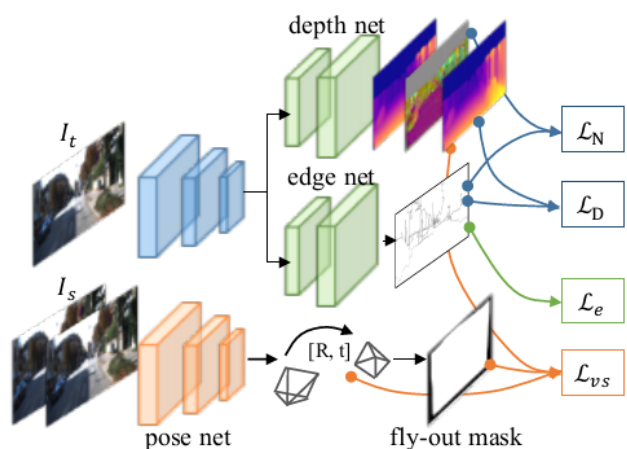


图 7. 同样为场景感知类任务, 以 3D-ASAP 的几何规则作为联系能得到有效的监督信号进行无监督多任务学习.

到场景的分割图, 并减少动态物体的影响, 如图 8

为了克服多任务下场景感知类任务例如深度估计等对于尺度的不确定性, [17] 首次提出基于速度的学习方法, 来加深模型对尺度的把握, 具体来说就是在位姿估计网络中, 将位置变换信息 t 提取出来, 基于此提取监督信号, 作为损失函数项, 促进位姿估计模型该方面的精度, 以此实现多任务下单个任务的优化训练, 如图 9 值的注意的是, 这里的速度信息是以有监督方式学习的, 这对模型中数据有着较高的要求, 为了克服此类需求, 也可以模仿 [14] 通过速度的前后一致性, 以无监督方式对速度学习进行优化.

3.4. 联邦学习范式

联邦学习本质上是一种分布式机器学习技术, 或机器学习框架. 联邦学习的目标是在保证数据隐私安全及合法合规的基础上, 实现共同建模, 提升 AI 模型的效果. 联邦学习最早在 2016 年由谷歌提出 [40], 原本用于解决安卓手机终端用户在本地更新模型的问题.

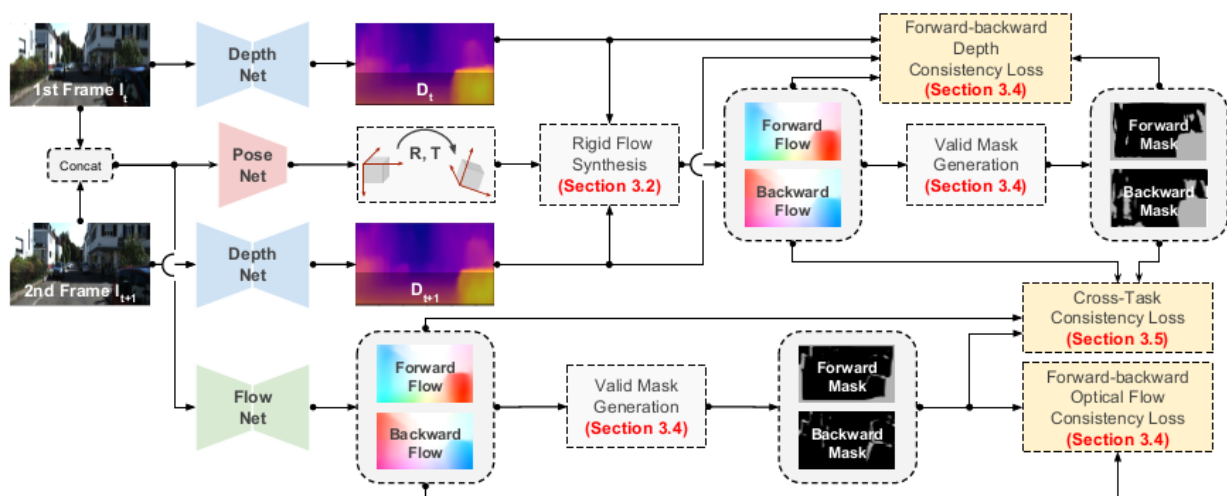


图 8. 通过场景前后向光流以及通过相机运动算出的场景光流进行两个任务的综合, 误差较大部分看做动态物体区域, 加强监督信号.

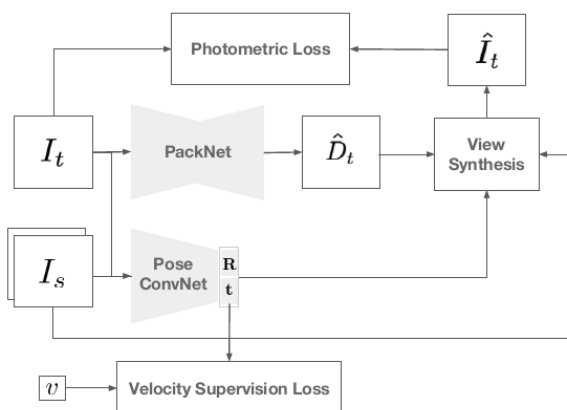


图 9. 通过模型对速度估计任务的学习, 模型能一定程度克服尺度歧义特点, 从而提高模型的鲁棒性

对于一些数据或者特殊场景下的应用来说, 其敏感程度并不能允许模型采用一般的训练, 而将模型简单的交付过去。例如在人脸识别应用中, 我们并不希望自己的人脸数据直接传输到服务商的存储器上并来训练模型, 另外, 在机场安全应用中, 有些军事机场规模较小数据量少, 飞机型号敏感, 此类数据更是关系到国家安全, 对待的态度尤其需要慎重。以上应用, 就涉及到多个同样任务的联合训练, 即联邦学习。

把每个参与共同建模的企业称为参与方, 根据多参与方之间数据分布的不同, 把联邦学习分为三类: 横向联邦学习、纵向联邦学习和联邦迁移学习¹。横

向联邦学习的本质是样本的联合, 适用于参与者间业态相同但触达客户不同, 即特征重叠多, 用户重叠少时的场景, 比如不同地区的银行间, 他们的业务相似 (特征相似), 但用户不同 (样本不同)。在传统的机器学习建模中, 通常是把模型训练需要的数据集合到一个数据中心然后再训练模型, 之后预测。在横向联邦学习中, 可以看作是基于一组样本的分布式模型训练, 分发全部数据到不同的机器, 每台机器从服务器下载模型, 然后利用本地数据训练模型, 之后返回给服务器需要更新的参数; 服务器聚合各机器上的返回的参数, 更新模型, 再把最新的模型反馈到每台机器。

在这个过程中, 每台机器下都是相同且完整的模型, 且机器之间不交流不依赖, 在预测时每台机器也可以独立预测, 可以把这个过程看成基于一组样本的分布式模型训练。谷歌最初就是采用横向联邦的方式解决安卓手机终端用户在本地更新模型的问题的纵向联邦学习的本质是特征的联合, 适用于用户重叠多, 特征重叠少的场景, 比如同一地区的商超和银行, 他们触达的用户都为该地区的居民 (样本相同), 但业务不同 (特征不同)。当参与者间特征和样本重叠都很少时可以考虑使用联邦迁移学习, 如不同地区的银行和商超间的联合。主要适用于以深度神经网络为基模型的场景。

在以上联邦学习中, 要针对采取的应用特点而应用不同策略。

3.5. 多任务与其他范式关系总结

多任务学习相关的范式有很多, 但并不是每个都能较好的应用在视觉任务中, 尤其是场景监测类任务集的衍生应用中, 例如机场安全、人脸识别等。

¹https://zhuanlan.zhihu.com/p/79284686?utm_source=wechat_session

在迁移学习范式下，多数目标任务的调优，又或者称为目标任务迁移，是在一个确定目标域上进行的，但是模型的初始化训练任务则极为固定，一般都是在ImageNet[41]上进行图像分类任务，以此作为初始化。此种方法可行性也有充足的工作来证明[42, 43]，而且在场景监测类任务中图像分类也是作为基础任务，特征空间应该与其余任务高度重合。但是，对于一些场景感知类任务，使用图像分类作为基础任务进行初始化虽能起到一定程度，但是任务之间的距离相对来说并不是那么近，因此针对场景感知类任务集，以场景感知类任务在合适的数据集上初始化[31]无疑能更快的使得模型在目标任务中收敛，因此对目标域能指向性灵活变换的源域开发是一个理想的方法。

根据数据的条件以及现有的资源，模型训练应该灵活调整是否我有监督、半监督还是无监督。例如在机场环境由虚拟工具搭建的情况下，通过有监督训练一个初始化模型就是一个合适的选择。在应用落地时，根据不同机场的环境以及规模等，可以将预训练模型在迁移学习的调优过程中更改监督方法。

由于数据的敏感以及信息安全等要求，联邦学习被提出。这中学习范式尤其适用于对安全领域要求较高的应用例如人脸识别以及机场安全。针对不同的规模以及性质的机场，联合了联邦学习范式的多任务学习模型应该以安全训练为前提，解决机场的运营安全问题。

4. 多任务系统中的网络结构设计问题

多任务学习是一种联合多个任务同时学习来增强模型表示和泛化能力的一种手段[44]，目前大都通过参数共享来实现多任务学习[45]。对于多任务学习的神经网络结构设计部分，很多工作都集中在寻找更好的神经网络参数共享机制上[46]，而好的神经网络参数共享机制依赖于多任务学习任务集中不同任务之间的关系。正如本文第二章提到的，机场场面态势感知任务主要分为场景感知任务和场景检测任务两类，本文在图10和图11中对这两类任务集中的任务关系进行了一个全面的梳理。就多任务系统中的网络结构设计方面，目前已有的基于深度学习的多任务学习工作提出了很多参数共享的策略，其中使用的较多的有硬共享，软共享，分层共享，另外还有一些比较新颖的值得探索的共享机制，比如梯度共享，元共享等[47][48][49][50][51]。本章从参数共享机制的角度对现有最常用的多任务深度神经网络模型进行了梳理，总结了现有的多任务深度神经网络结构设计中存在的关键问题，最后针对这些关键问题提出了一些优化的策略。

4.1. 多任务深度神经网络结构

4.1.1 基于参数硬共享的多任务深度神经网络结构

参数硬共享是目前应用最为广泛的共享机制，它不同任务底层共享模型结构和参数，顶层分为几个不同的目标进行网络训练，这种结构本质上可以减少过拟合的风险，但是效果上可能受到任务差异和数据分

布带来的影响。基本上，只要是能预测单模型的模型，都可以很简单的转化为参数硬共享模型的结构，只需要将共享层的最后一层与多个输出层拼接即可。参数硬共享模型的原理如图10所示。

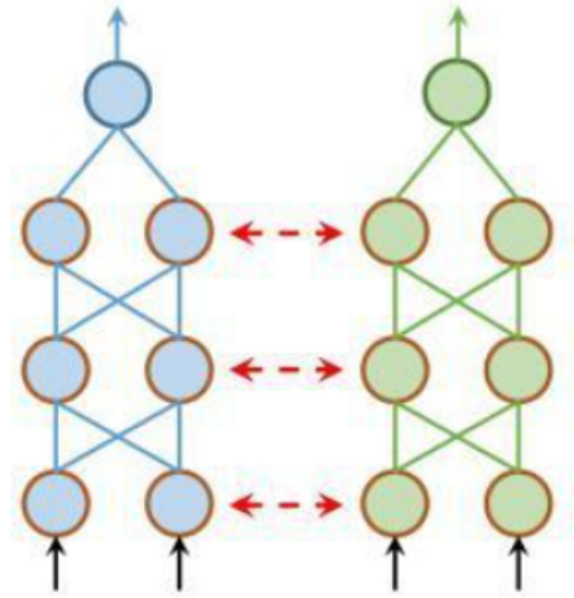


图 10. 参数硬共享示意图

接下来介绍两个典型的基于参数硬共享的多任务深度神经网络模型：

(1) MMoE多任务网络模型：MMoE模型的结构基于广泛使用的Shared-Bottom结构和MoE结构[52][53]，如图11所示，其中(a)是传统的硬共享参数模型，(b)是MoE模型，使用单个gate控制多个任务的参数，(c)是MMoE模型，在MoE的基础上，每个任务使用一个gate控制其权重。11中的expert指对模型输入进行不同方式的变换处理的网络层，每个Expert表示一种网络（Expert也可以都一样）。gate控制每个Expert权重的变量，对于每一个任务，不同Expert的权重可能是不一样的，因此使用gate来控制权重，类似于注意力机制[54]。

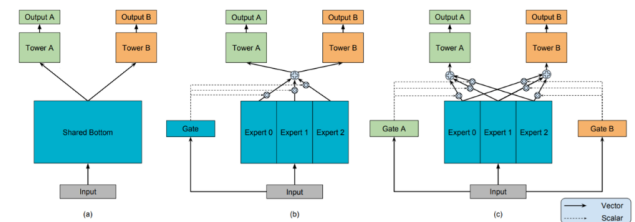


图 11. MMoE网络结构示意图：a)共享浅层模型；b)单门MoE模型；c)多门MoE模型

MoE模型对于不同的任务的gate权重是一样的，其

函数表达式如公式1所示，其中k表示第k个任务，n表示n个expert网络。

$$y^k = h^k \sum_{i=1}^n g_i f_i(x) \quad (1)$$

MMoE是在MoE的基础上提出的方法，作者认为对于不同的任务，模型的权重选择是不同的，所以为每个任务分配一个gate模型。对于不同的任务，gate k的输出表示不同Expert被选择的概率，将多个Expert加权求和，得到 $f_k(x)$ ，并输出给特点的Tower模型，用于最终的输出。MMoE模型的表达式如公式2所示：

$$[f^k(x) = \sum_{i=1}^n g_i^k(x) f_i(x)] \quad (2)$$

其中：

$$[g^k(x) = \text{soft max}(W_{g^k(x)})] \quad (3)$$

类似于MoE，k表示第k个任务，每个任务对应一个gate。MMoE的底层参数仍然是共享的，但是通过目标和网络参数直接的gate来学习，让每部分网络充分学习到对每个目标的贡献最大的一组参数结构，通过这种方式来保证底层网络参数共享的时候，不会出现目标之间相互抵消的作用。MMoE模型从一定程度上解决了多个目标（任务）在训练过程中的相互耦合的问题，即使用门控概念降低了因为共享网络部分带来的特征耦合。但其实这是不够的，因为在每一个expert内部，与其他expert不存在联系，这导致每个expert的表达能力不是很强[55]。

(2)SNR多任务网络模型：当任务之间相关性比较强的时候，MTL可以学习到多个任务之间的关系，但是当任务之间相关性比较差的时候，在预估准确度上会比较差，主要是任务之间的干扰造成的共享网络部分难以收敛，MTL需要人工去调整模型的结构，以前的参数软共享在灵活性上和计算复杂度上不能兼得。同时，为了解决MMoE模型的局限性，SNR多任务学习模型被提出。该模型的作者提出了灵活参数共享的概念，即我们不应把共享网络部分作为整体的参数分享给每一个需要训练的目标，在共享网络的内部也需要互相共享参数，以提高表达[56]。通过设计了一款Sub-Network Routing模型将共享网络的上下层进行剥离，用下层中的所有参数作为上层输入共享。SNR多任务学习模型对任务之间的相关性强弱不敏感，借助简单的NAS（Neural Architecture Search）[57][58]，可以对自网络进行组合，学习一个好的模型结构。其中，该模型的作者设计了两种共享方式：transformation和average，具体的网络结构如图12所示。

SNR多任务学习模型的主要优势是把网络结构的跨任务参数共享抽象为网络子结构的路由问题；引入0-1隐变量对路由作最优化；通过L1正则化（可以得

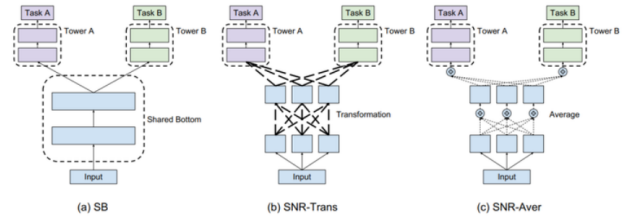


图 12. SNR网络结构示意图：a)SB模型；b)SNR-Trans模型；c)SNR-Aver模型

到参数量更小的网络），学到稀疏解，同等精度下节约11%的网络结构开销[59]。

4.1.2 基于参数软共享的多任务深度神经网络结构

参数软共享不同于参数硬共享，每个任务有自己的参数，最后通过对不同任务的参数之间的差异加约束，表达相似性，比如可以使用L2正则化和迹范数等加以约束[60][61][62]。参数软共享模型的原理如图13所示。

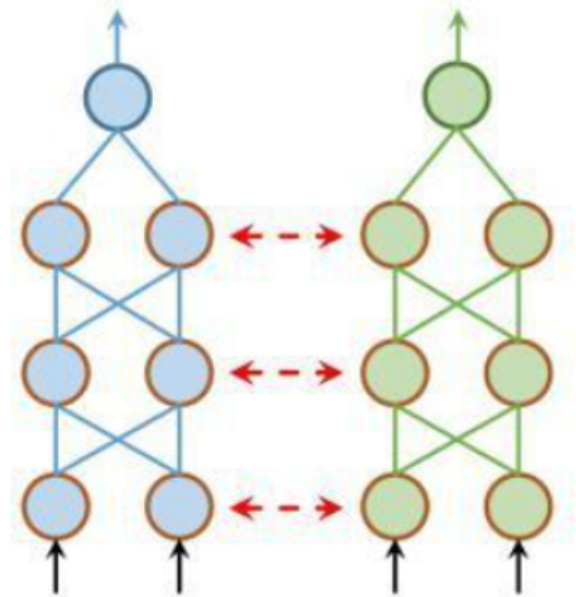


图 13. 参数软共享示意图

接下来介绍一个典型的基于参数软共享的多任务深度神经网络模型：MTAN多任务网络模型。MTAN结构主要包括两大部分，一个任务共享的主网络和K个特定任务的子网络，共享网络可以根据特定的任务进行设计，而每个特定于任务的子网络由一组注意力模块组成，这些模块与共享网络相连接。每个注意力模块对共享网络的特定层应用一个软注意力mask，以学习特定于任务的特征。基于这种设计，共享主网络可以

看做是一个跨任务的特征表示，每一个注意力mask都可以被看作是对主网络的特征选择器，决定哪些共享特征被用到自己的子任务中去[63]。14是MTAN的结构图，以VGG-16为主网络架构，因为SegNet的对称设计，只给出了编码器部分，任务一（绿色）和任务二（蓝色）的注意力模块，与共享网络（灰色）相连，决定了主网络那些特征会被利用到子网络中。中间是一个注意力模块的内部结构。

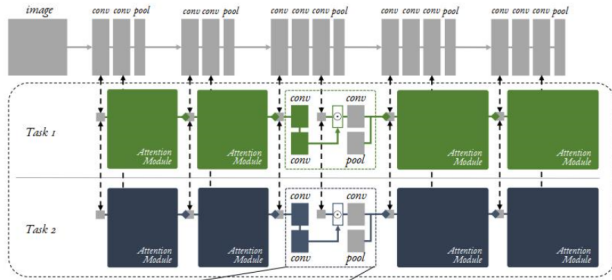


图 14. MTAN网络结构示意图

我们把注意力模块的内部结构再细化，如下图所示，该层的共享特（U）首先跟上一个注意力模块的输出进行Merge，具体方式为Concatenation(维度相加)，然后，将融合后的结果作为输入，经过下面这些操作：1) g : 1×1 卷积，BN层，ReLU激活函数 2) h : 1×1 卷积，BN层，Sigmoid激活函数 3) a : 得到对应的注意力mask 4) p : 将注意力mask与主网络特征进行Element-wise Multiplication（对应元素相乘） 5) a : 得到筛选出的特征 6) f : 3×3 卷积，BN层，ReLU激活函数

如图15所示，由于每个注意力模块中的mask是根据主网络相应层的共享特征学到的，因此两者可以联合学习，使得网络在不同任务中泛化性能更好，同时提升在特定任务中的表现。MTAN网络体系结构由一个全局特性池和特定任务的注意力模块组成，允许以端到端的方式自动学习任务共享和特定于任务的特性[33]。

4.2. 网络结构设计中的问题

多任务网络模型的性能与不同任务的信息融合效率密切相关，基于参数硬共享的多任务深度神经网络模型把多个任务的数据表示嵌入到同一个语义空间中，再为每个任务使用单一任务特定层提取任务特定表示[64][65]。基于参数硬共享的多任务深度神经网络模型适合处理有较强相关性的任务，但遇到弱相关任务时常常表现很差[66][67][68]。而基于参数软共享的多任务深度神经网络模型为每个任务都学习一个网络，但每个任务的网络都可以访问其他任务对应网络中的信息，例如表示、梯度等。参数软共享机制非常灵活，不需要对任务相关性做任何假设，但是由于为每个任务分配一个网络，常常需要增加很多参数[69][70]，而且现有基于参数软共享的多任务深度神经网络模型中不同任务的信息交互层面相对单一，通常只是对同一

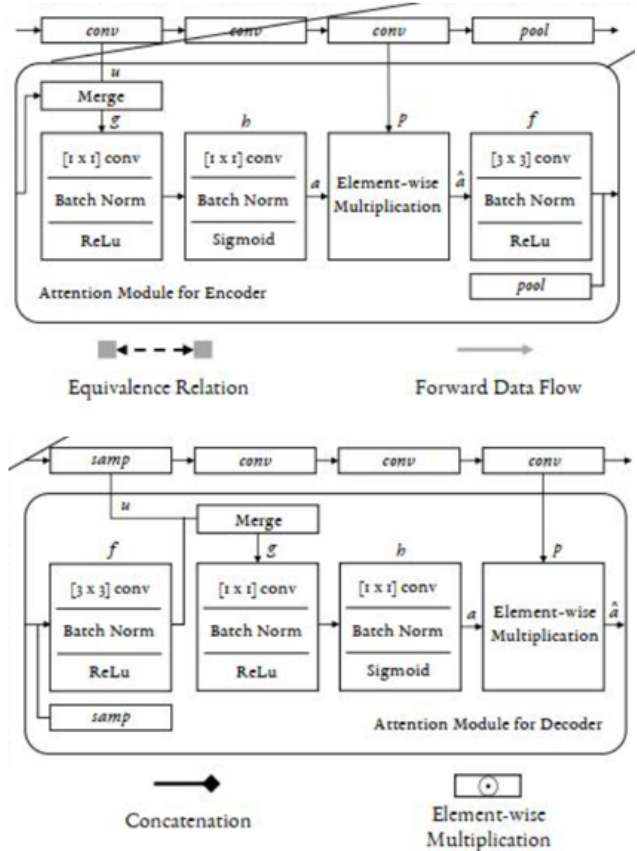


图 15. MTAN网络中的注意力模块示意图

层网络信息进行交互，没有考虑不同任务间可能存在的尺度差异问题，例如机场场面态势感知中的检测任务和实例分割任务。

4.3. 优化策略

针对上一节提到的多任务网络结构设计中面临的两个问题，本文提出了两种具有针对性的优化策略。其中，针对基于参数硬共享的多任务深度神经网络模型中弱相关任务表现差的问题，本文提出了一种基于任务路由机制的动态多任务深度神经网络结构进行优化；而针对基于参数软共享的多任务深度神经网络模型中不同任务信息交互层面单一问题，本文提出了一种多尺度信息蒸馏交互深度神经网络结构进行优化。

(1) 基于任务路由机制的动态多任务深度神经网络结构：通过在网络结构中引入任务路由层，它可以将任务特定的二进制掩码应用于共享网络部分的输出，归零化不适用于特定任务的特征，并有效地为各个任务分配一个与其他任务不同的子网络[58]。在[58]的基础上，将二进制0, 1掩码修改为类似softmax操作的带权重特征融合掩码，同时实现对不适用特征的归零化和适用特征的有效融合。

(2) 多尺度信息蒸馏交互深度神经网络结构：在多任务学习设置中，提取任务信息时在多个尺度上考

虑任务交互的重要性,在某个尺度上具有高亲和力的任务不能保证在其他尺度上保留这种行为,反之亦然。因此,可以考虑首先通过多尺度多模态蒸馏单元[71]在每个尺度上显式地模拟任务交互。其次,通过特征传播模块将提取的任务信息从较低的尺度传播到较高的尺度。最后,通过一个特征聚合单元聚合来自所有尺度的细化任务特征,以产生最终的各个任务的预测结果。

4.4. 总结

机场场面多任务态势感知问题主要分为场景感知任务和场景检测任务两类,其中场景感知任务通常为密集估计任务,而场景检测任务通常为非密集估计任务。场景感知任务和场景检测任务间的相关性通常较弱,因此采用基于任务路由机制的动态多任务神经网络结构可以有效地避免弱相关性任务在训练过程中的相互干扰问题。且在机场场面多任务中,不同类型的任务之间信息传递的方式应不局限在同一个尺度上,因此,多尺度信息蒸馏交互神经网络结构有助于上述场景感知类任务和场景检测类任务的特征融合。

参考文献

- [1] Rich Caruana. Multitask learning. *Machine learning*, 28(1):41–75, 1997. 1
- [2] Sinno Jialin Pan and Qiang Yang. A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, 22(10):1345–1359, 2009. 1
- [3] German I Parisi, Ronald Kemker, Jose L Part, Christopher Kanan, and Stefan Wermter. Continual lifelong learning with neural networks: A review. *Neural Networks*, 113:54–71, 2019. 1
- [4] Matteo Hessel, Hubert Soyer, Lasse Espeholt, Wojciech Czarnecki, Simon Schmitt, and Hado van Hasselt. Multi-task deep reinforcement learning with popart. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 3796–3803, 2019. 1
- [5] Iasonas Kokkinos. Ubernet: Training a universal convolutional neural network for low-, mid-, and high-level vision using diverse datasets and limited memory. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 6129–6138, 2017. 2
- [6] Kaipeng Zhang, Zhanpeng Zhang, Zhifeng Li, and Yu Qiao. Joint face detection and alignment using multitask cascaded convolutional networks. *IEEE Signal Processing Letters*, 23(10):1499–1503, 2016. 2
- [7] Anurag Ranjan, Varun Jampani, Lukas Balles, Kihwan Kim, Deqing Sun, Jonas Wulff, and Michael J Black. Competitive collaboration: Joint unsupervised learning of depth, camera motion, optical flow and motion segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12240–12249, 2019. 2
- [8] Eigen. Depth map prediction from a single image using a multi-scale deep network. In *NIPS*, 2014. 2
- [9] David Eigen and Rob Fergus. Predicting depth, surface normals and semantic labels with a common multi-scale convolutional architecture. In *Proceedings of the IEEE international conference on computer vision*, pages 2650–2658, 2015. 2
- [10] Wei Yin, Yifan Liu, Chunhua Shen, and Youliang Yan. Enforcing geometric constraints of virtual normal for depth prediction. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 5684–5693, 2019. 2
- [11] Zhenheng Yang, Peng Wang, Yang Wang, Wei Xu, and Ram Nevatia. Lego: Learning edge with geometry all at once by watching videos. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 225–234, 2018. 2, 4
- [12] Alexey Dosovitskiy, Philipp Fischer, Eddy Ilg, Philip Hausser, Caner Hazirbas, Vladimir Golkov, Patrick Van Der Smagt, Daniel Cremers, and Thomas Brox. Flownet: Learning optical flow with convolutional networks. In *Proceedings of the IEEE international conference on computer vision*, pages 2758–2766, 2015. 2
- [13] Zhichao Yin and Jianping Shi. Geonet: Unsupervised learning of dense depth, optical flow and camera pose. In *CVPR*, 2018. 2
- [14] Yuliang Zou, Zelun Luo, and Jia Bin Huang. DF-Net: Unsupervised Joint Learning of Depth and Flow Using Cross-Task Consistency. In *ECCV*, 2018. 2, 4
- [15] Marvin Klingner, Jan-Aike Termohlen, Jonas Mikolajczyk, and Tim Fingscheidt. Self-supervised monocular depth estimation: Solving the dynamic object problem by semantic guidance. In *European Conference on Computer Vision*, pages 582–600. Springer, 2020. 2
- [16] Tinghui Zhou, Matthew Brown, Noah Snavely, and David G. Lowe. Unsupervised learning of depth and ego-motion from video. In *CVPR*, 2017. 2
- [17] Vitor Guizilini, Rares Ambrus, Sudeep Pillai, Allan Raventos, and Adrien Gaidon. 3d packing for self-supervised monocular depth estimation. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020. 2, 4
- [18] Amir R Zamir, Alexander Sax, William Shen, Leonidas J Guibas, Jitendra Malik, and Silvio Savarese. Taskonomy: Disentangling task transfer learning. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3712–3722, 2018. 2
- [19] Trevor Standley, Amir Zamir, Dawn Chen, Leonidas Guibas, Jitendra Malik, and Silvio Savarese. Which tasks should be learned together in multi-task learning. In *International Conference on Machine Learning*, pages 9120–9132, 2020. 2
- [20] Amir R Zamir, Alexander Sax, Nikhil Cheerla, Rohan Suri, Zhangjie Cao, Jitendra Malik, and Leonidas J Guibas. Robust learning through cross-task consistency. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11197–11206, 2020. 2
- [21] Rajeev Ranjan, Vishal M Patel, and Rama Chellappa. Hyperface: A deep multi-task learning framework for face detection, landmark localization, pose estimation, and gender

- recognition. *IEEE transactions on pattern analysis and machine intelligence*, 41(1):121–135, 2017. 3
- [22] Ravi Garg, Vijay Kumar Bg, Gustavo Carneiro, and Ian Reid. ‘geometry to the rescue’. In *European conference on computer vision*, pages 740–756. Springer, 2016. 3
- [23] John Flynn, Ivan Neulander, James Philbin, and Noah Snavely. Deepstereo: Learning to predict new views from the world’s imagery. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5515–5524, 2016. 3
- [24] Qiang Yang, Yu Zhang, Wenyuan Dai, and Sinno Jialin Pan. *Transfer Learning*. Cambridge University Press, 2020. 3
- [25] Min-Ling Zhang and Zhi-Hua Zhou. A review on multi-label learning algorithms. *IEEE TKDE*, 2014. 3
- [26] German Ignacio Parisi, Ronald Kemker, Jose L. Part, Christopher Kanan, and Stefan Wermter. Continual lifelong learning with neural networks: A review. *Neural Networks*, 113:54–71, 2019. 3
- [27] A. Gaidon, Q. Wang, Y. Cabon, and E. Vig. Virtualworlds as proxy for multi-object tracking analysis. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016. 3
- [28] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun. Vision meets robotics: The kitti dataset. *IJRR*, 2013. 3
- [29] Jonathan Tremblay, Aayush Prakash, David Acuna, Mark Brophy, Varun Jampani, Cem Anil, Thang To, Eric Cameracci, Shaaad Boochoon, and Stan Birchfield. Training deep networks with synthetic data: Bridging the reality gap by domain randomization. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 969–977, 2018. 3
- [30] Josh Tobin, Rachel Fong, Alex Ray, Jonas Schneider, Wojciech Zaremba, and Pieter Abbeel. Domain randomization for transferring deep neural networks from simulation to the real world. In *2017 IEEE/RSJ international conference on intelligent robots and systems (IROS)*, pages 23–30. IEEE, 2017. 3
- [31] Kaixuan Wang, Yao Chen, Hengkai Guo, Linfu Wen, and Shaojie Shen. Geometric pretraining for monocular depth estimation. In *2020 IEEE International Conference on Robotics and Automation (ICRA)*, pages 4782–4788. IEEE, 2020. 3, 6
- [32] Qiuhua Liu, Xuejun Liao, and Lawrence Carin. Semi-supervised multitask learning. In *NIPS*, 2007. 4
- [33] Qiuhua Liu, Xuejun Liao, Hui Li, Jason R. Stack, and Lawrence Carin. Semisupervised multitask learning. *IEEE TPAMI*, 2009. 4
- [34] Yu Zhang and Dit-Yan Yeung. Semi-supervised multi-task regression. In *ECMLPKDD*, 2009. 4
- [35] Roi Reichart, Katrin Tomanek, Udo Hahn, and Ari Rapoport. Multi-task active learning for linguistic annotations. In *ACL*, 2008. 4
- [36] Ayan Acharya, Raymond J. Mooney, and Joydeep Ghosh. Active multitask learning using both latent and supervised shared topics. In *SDM*, 2014. 4
- [37] Meng Fang and Dacheng Tao. Active multi-task learning via bandits. In *SDM*, 2015. 4
- [38] Hui Li, Xuejun Liao, and Lawrence Carin. Active learning for semi-supervised multi-task learning. In *ICASSP*, 2009. 4
- [39] Anastasia Pentina and Christoph H. Lampert. Multi-task learning with labeled and unlabeled tasks. In *ICML*, 2017. 4
- [40] Jakub Konečný, H Brendan McMahan, Felix X Yu, Peter Richtárik, Ananda Theertha Suresh, and Dave Bacon. Federated learning: Strategies for improving communication efficiency. *arXiv preprint arXiv:1610.05492*, 2016. 4
- [41] Deng Jia and Wei Dong. Imagenet: A large-scale hierarchical image database. In *CVPR*, 2009. 6
- [42] Dan Hendrycks, Kimin Lee, and Mantas Mazeika. Using pre-training can improve model robustness and uncertainty. In *International Conference on Machine Learning*, pages 2712–2721. PMLR, 2019. 6
- [43] Kaiming He, Ross Girshick, and Piotr Dollár. Rethinking imagenet pre-training. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4918–4927, 2019. 6
- [44] Shikun Liu, Edward Johns, and Andrew J Davison. End-to-end multi-task learning with attention. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1871–1880, 2019. 6
- [45] Jiaqi Ma, Zhe Zhao, Jilin Chen, Ang Li, Lichan Hong, and Ed H Chi. Snr: Sub-network routing for flexible parameter sharing in multi-task learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 216–223, 2019. 6
- [46] Jiaqi Ma, Zhe Zhao, Xinyang Yi, Jilin Chen, Lichan Hong, and Ed H Chi. Modeling task relationships in multi-task learning with multi-gate mixture-of-experts. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 1930–1939, 2018. 6
- [47] Tianxiang Sun, Yunfan Shao, Xiaonan Li, Pengfei Liu, Hang Yan, Xipeng Qiu, and Xuanjing Huang. Learning sparse sharing architectures for multiple tasks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 8936–8943, 2020. 6
- [48] Sebastian Ruder. An overview of multi-task learning in deep neural networks. *arXiv preprint arXiv:1706.05098*, 2017. 6
- [49] Yu Zhang and Qiang Yang. An overview of multi-task learning. *National Science Review*, 5(1):30–43, 2018. 6
- [50] Alex Kendall, Yarin Gal, and Roberto Cipolla. Multi-task learning using uncertainty to weigh losses for scene geometry and semantics. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7482–7491, 2018. 6
- [51] Varun Ravi Kumar, Senthil Yogamani, Hazem Rashed, Ganesh Sitsu, Christian Witt, Isabelle Leang, Stefan Milz, and Patrick Mäder. Omnidet: Surround view cameras based

- multi-task visual perception network for autonomous driving. *IEEE Robotics and Automation Letters*, 6(2):2830–2837, 2021. 6
- [52] Xi Lin, Hui-Ling Zhen, Zhenhua Li, Qing-Fu Zhang, and Sam Kwong. Pareto multi-task learning. *Advances in neural information processing systems*, 32:12060–12070, 2019. 6
- [53] Arun Mallya, Dillon Davis, and Svetlana Lazebnik. Piggyback: Adapting a single network to multiple tasks by learning to mask weights. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 67–82, 2018. 6
- [54] Kevis-Kokitsi Maninis, Ilija Radosavovic, and Iasonas Kokkinos. Attentive single-tasking of multiple tasks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1851–1860, 2019. 6
- [55] Ishan Misra, Abhinav Shrivastava, Abhinav Gupta, and Martial Hebert. Cross-stitch networks for multi-task learning. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3994–4003, 2016. 7
- [56] Sylvestre-Alvise Rebuffi, Hakan Bilen, and Andrea Vedaldi. Efficient parametrization of multi-domain deep neural networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 8119–8127, 2018. 7
- [57] Sebastian Ruder, Joachim Bingel, Isabelle Augenstein, and Anders Søgaard. Latent multi-task architecture learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 4822–4829, 2019. 7
- [58] Gjorgji Strezoski, Nanne van Noord, and Marcel Worring. Many task learning with task routing. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 1375–1384, 2019. 7, 8
- [59] Gjorgji Strezoski, Nanne van Noord, and Marcel Worring. Learning task relatedness in multi-task learning for images in context. In *Proceedings of the 2019 on International Conference on Multimedia Retrieval*, pages 78–86, 2019. 7
- [60] Sumanth Chennupati, Ganesh Sistu, Senthil Yogamani, and Samir A Rawashdeh. Multinet++: Multi-stream feature aggregation and geometric loss strategy for multi-task learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 0–0, 2019. 7
- [61] Jifeng Dai, Kaiming He, and Jian Sun. Instance-aware semantic segmentation via multi-task network cascades. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3150–3158, 2016. 7
- [62] Harm De Vries, Florian Strub, Sarath Chandar, Olivier Pietquin, Hugo Larochelle, and Aaron Courville. Guess-what?! visual object discovery through multi-modal dialogue. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5503–5512, 2017. 7
- [63] Dan Xu, Wanli Ouyang, Xiaogang Wang, and Nicu Sebe. Pad-net: Multi-tasks guided prediction-and-distillation network for simultaneous depth estimation and scene parsing. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 675–684, 2018. 8
- [64] Ozan Sener and Vladlen Koltun. Multi-task learning as multi-objective optimization. *arXiv preprint arXiv:1810.04650*, 2018. 8
- [65] Xi Lin, Zhiyuan Yang, Qingfu Zhang, and Sam Kwong. Controllable pareto multi-task learning. *arXiv preprint arXiv:2010.06313*, 2020. 8
- [66] Virginia Smith, Chao-Kai Chiang, Maziar Sanjabi, and Ameet Talwalkar. Federated multi-task learning. *arXiv preprint arXiv:1705.10467*, 2017. 8
- [67] Tadas Baltrušaitis, Chaitanya Ahuja, and Louis-Philippe Morency. Multimodal machine learning: A survey and taxonomy. *IEEE transactions on pattern analysis and machine intelligence*, 41(2):423–443, 2018. 8
- [68] Chanho Ahn, Eunwoo Kim, and Songhwai Oh. Deep elastic networks with model selection for multi-task learning. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 6529–6538, 2019. 8
- [69] Carlo D’Eramo, Davide Tateo, Andrea Bonarini, Marcello Restelli, and Jan Peters. Sharing knowledge in multi-task deep reinforcement learning. In *International Conference on Learning Representations*, 2019. 8
- [70] Carl Doersch and Andrew Zisserman. Multi-task self-supervised visual learning. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2051–2060, 2017. 8
- [71] Simon Vandenhende, Stamatios Georgoulis, and Luc Van Gool. Mti-net: Multi-scale task interaction networks for multi-task learning. In *European Conference on Computer Vision*, pages 527–543. Springer, 2020. 9