1. 人体模型

人体模型的构建和描述是人体姿态检测流程中一个重要的模块。当前主流的人体模型主要有运动模型（Kinematic Model）、平面模型（Planar Model）和体积模型（Volumetric Model）­­[1]。

* 1. 运动模型

运动模型利用关节点的位置和身体部位的方向来定义人体姿态。运动模型又可以分为预定义模型（predefined model）和图形结构模型（graph structure）。其中，PSM（Pictorial Structures Model）[2,3]是一种常用的图形结构模型，也是当前人体姿态检测领域里相当重要的里程碑。Pictorial Structures[4]早由Fischler 和 Elschlager 于1973年提出，这种方法将人体姿态检测问题转化为能量方程的最小化问题。后来Felzenszwalb [3]人在此模型的基础上定义了人体姿态。Felzenszwalb的方法虽然只能应用于前景背景分离的灰度图像上，但仍为后续研究成果的改进和实现提供了诸多帮助。

树结构模型（tree-structured model）是PSM的一个特例，无论在2D还是3D上，都适用于人体姿态检测[5–13]。然而运动树结构模型在表示身体部位之间的依赖关系时有一定的局限性。此外，当遮挡存在的情况下，身体部位完全无法检测[14]。针对运动树结构模型的局限性，后续许多研究提出在树结构模型上的改进方法。Wang等人[15]通过添加不同形状的树结构模型来解决缺少模型描述符的问题。Sapp等人 [16]通过添加模型的状态来增强模型的表达能力。Wang等人[15]和Komodakis等人[17] 使用多个树模型而不是单个树模型进行人类姿态估计。

* 1. 平面模型

平面模型不仅能够捕获身体部位之间的连接关系，还能够学习人体部位的形状和外观。 ASM（Active Shape Models）就是一种平面模型，能够用于表示整个人体结构，并用主成分分析（PCA）从平均形状中获得轮廓变形的统计[18,19]。纸板模型（Cardboard Model）是另一种平面模型，它由前景物体颜色信息和人体部位矩形构成。纸板模型通常包含一个躯干和八个半截肢体，前景信息由颜色直方图表示，每个身体部位由该部位的平均RGB来表示。例如Hao等人[20]的方法就是采用纸板模型来实现人体姿态检测。

* 1. 体积模型

体积模型能够真实地表示3D体形和姿态。几何形状和网格是体积模型中常用的两种方式。当使用几何形状作为模型组件时，人体部分用圆柱体、圆锥体和其他近似形状来表示肢体，例如Hedvig等人[21]的方法就是用圆柱体的复合体建模人体姿态。 这种模型能够准确地获得人体四肢的真实形状，并且容许人体结构或服饰的变化，因而比图形基于结构的方法更具有准确性。当使用网格来表示人体姿态时，网格通常被分割成几个身体部位。SCAPE（Shape Completion and Animation of People）是一种常用的3D网格模型[22,23]。

* 1. 姿势先验

人体姿势受到许多因素的约束，包括运动学、关节操作限制以及特定活动下的运动模式等[24]。通过从数据中学习姿势约束并创建姿势先验，可以为人体姿态检测提供有效的信息。一些研究将这些学习到的知识作为模型的先验知识并用于约束人体关节点[25] 。

1. 分类算法

通常情况下，人体姿态的检测的分类算法可划分为两类：产生式分类算法和判别式分类算法。产生式分类算法把人体姿态检测模拟为几何映射问题，判别式分类算法则把人体姿态检测视为特定图像处理问题。另一种划分方法根据起始点为高级抽象语义或低级像素，把人体姿态的检测方法划分为自顶向下或自底向上。

* 1. 产生式分类和判别式分类
     1. 产生式分类算法

产生式分类算法的基本思想是首先建立类别样本的概率密度模型，再利用模型进行推理预测，这种方法要求具有大量的已知样本。在人体姿态检测问题中，产生式分类算法[26]从人体姿势初始化模型开始，将姿势映射到图像平面，再用图像上的信息验证模型的准确性。

* + 1. 判别式分类算法

判别式模型与生成式模型的工作原理恰好相反，它的的基本思想是通过大量的样本学习获得模型分布。在人体姿态检测问题中，判别式分类算法从图像信息开始，利用训练数据为图像信息与人体姿态之间的关系建模。这种方法相比于产生式分类算法相对简单、计算效率高，一旦建模完成，对输入图像的姿态检测效率远远高于产生式分类算法。大部分的人体姿态检测方法采用的是判别式的分类算法。判别式分类算法又可以根据其训练集的来源分为基于学习的分类算法和基于特定样本的分类算法。

基于学习的分类算法不对训练样本进行特定筛选和分类，现已存在大量这一类方法相关的文献。SVM（Support Vector Machines）是当前被最广泛使用的分类器。Remi等人[27]用SVM或相关性向量机（RVM）学习获得的专用检测器替代简单的部位检测器。Ryuzo等人[28]为了区分姿态簇，使用具有姿势相关特征选择的内核支持向量机，通过最小化半径/边界约束来估计RBF内核的尺度参数，从而实现了内核支持向量机的特征选择。Huazhong等人[29]使用的判别式词袋方法实现单目图像的3D人体姿态检测。近年来，随着深度学习的快速发展及其在数据挖掘和表达能力上的优势，人体姿态检测领域引入了深度学习方法用于学习图像中的关键信息，其中CNN（Convolutional Neural Networks）被广泛应用于人体姿态检测中。Tompson等人[30]提出了一种涉及深卷积网络和马尔可夫随机场（MRF）模型的组合架构。Gkioxari等人[31]用由CNN特征表示的区域来训练出具有损失函数的R-CNN检测器。Joao等人[32]采用迭代误差反馈，通过反馈误差预测来改变初始解。

基于特定样本的分类算法中，对于输入的图像，我们用特定姿态的离散图像集合及其对应的特征表示来训练模型并实现人体姿态检测[33]。随机树[34]和随机森林[35,36]、霍夫森林[37]都是可以用于快速处理这种类型问题强大分类技术[38]。此外，稀疏表示（SR）也常用于提取最重要的训练样本，后来所有的估计也都是基于这些样本进行的[39–41]。

* + 1. 融合产生式与判别式分类算法

产生式分类将人体模型投射到2D图像空间中，并测量它们之间的距离[33]，而判别式分类检测人体的各部位来重建人体姿势。由于产生式分类算法效率低下，而判别式算法对于不存在于训练集中的姿势识别率低 [29]，许多研究结合了这两种方法的优势，提出产生式分类与判别式分类相结合的分类算法。这些方法通常用判别式方法[42]的检测结果来初始化姿势，并用产生式方法来优化局部区域内的人体姿态[43–45]。

* 1. 自顶向下和自底向上

根据人体姿态检测算法的工作方式是开始于高级抽象语义或低级像素，我们又可以把人体姿态检测算法分为自顶向下和自底向上的方法。在计算机视觉的语义层次结构中，图像是最低级别，而人体姿态的配置以及人体姿态所属的行为类型是较高级别。

* + 1. 自底向上

自底向上的方法从图像中采集并描述出具有代表性的特征，将这些特征直接用语检测人体姿态或者用于定位身体部位再组合成人体姿态。根据单位面积的大小，这一类方法包括了基于像素超像素的方法和基于人体部位的方法。

1. 基于像素或超像素的方法

图像的像素或超像素可以用于分割图像并集成到姿态检测中 [46,47]。基于像素的方法也可以与其他方法相结合。例如，Hernández Vela等人[48]利用能量最小化框架，通过图切割优化来扩展每像素的分类方法。此外，还可以利用分割的结果来增强像素级的估计。 Marcin等人[47]提出了一种方法，通过在检测到的区域上初始化的“GrabCut”来进一步缩小搜索空间，直到缩小到身体部位的搜索空间[49,50]。基于部件和基于像素的方法也可以在单个优化框架中组合[51]。超像素也可用于限制人体模型中的关节位置[52]。在基于超像素的方法中，我们还可以优化身体部位匹配和前景估计。例如，利用分支绑定算法[53–55]来实现这个过程。

（2）基于部件的方法

基于部件的方法通过学习人体部位外观和位置模型来解决人体姿态检测问题。在基于部分的方法中，首先从图像中检测身体部位候选位置，然后逐个检测身体部位使其符合人体外观的正常表现[56]。Yang等人[12]扩展了可变形部件模型（DPM）[13,57]，提出了用于铰接式2D人体姿态检测的FMP（Flexible mixture of parts ）模型 。 这种方法后来利用组合和/或图形语法模型被进一步改进[58]。

* + 1. 自顶向下

我们这里的自顶向下指的是从高级抽象语义到低级像素的问题解决过程[59]，其中高级语义用于指导低级语义。根据这个概念，自顶向下的方法更多是结合自底向上的方法来使用，而且更高级的语义通常能在效果上提供很大帮助。

* + 1. 融合自顶向下与自底向上

自顶向下与自底向上的方法结合起来比产生式方法与判别式方法结合起来更加灵活。通过将图形运动学模型与检测方法相结合，可以同时获得检测结果和3D姿态[56,60,61]。此外，通过图形切割可以找到全局最小能量[62]。 Bray等人[60]使用图形切割来优化姿态参数以实体人体的集成分割和3D姿态估计。

1. Gong, W. *et al.* Human pose estimation from monocular images: A comprehensive survey. *Sensors (Switzerland)* **16,** 1–39 (2016).

2. Felzenszwalb, P. F., Huttenlocher, D. P. & Pedro F. Felzenszwalb. Efficient maching of pictorial structures. *Cvpr* 66–75 (2000).

3. Felzenszwalb, P. F. & Huttenlocher, D. Pictoral structures for object recognition. *Ijcv* **61,** 55–79 (2005).

4. Fischler, M. A. & Elschlager, R. A. The Representation and Matching of Pictorial Structures Representation. *IEEE Trans. Comput.* **C-22,** 67–92 (1973).

5. Andriluka, M., Roth, S. & Schiele, B. Pictorial structures revisited: People detection and articulated pose estimation. *2009 IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Work. CVPR Work. 2009* 1014–1021 (2009). doi:10.1109/CVPRW.2009.5206754

6. Felzenszwalb, P., McAllester, D. & Ramanan, D. A Discriminatively Trained, Multiscaled, Deformable Part Model. *Cvpr* 1–8 (2008). doi:10.1109/CVPR.2008.4587597

7. Andriluka, M., Roth, S. & Schiele, B. Monocular 3D Pose Estimation and Tracking by Detection. *2010 Ieee Conf. Comput. Vis. Pattern Recognit.* 623–630 (2010). doi:10.1109/CVPR.2010.5540156

8. Yang, Y. & Ramanan, D. Articulated pose estimation with ﬂexible mixtures-of-parts. *Comput. Vis. Pattern …* (2011).

9. Sapp, B., Jordan, C. & Taskar, B. Adaptive pose priors for pictorial structures. *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.* 422–429 (2010). doi:10.1109/CVPR.2010.5540182

10. Chen, X. & Yuille, A. Parsing occluded people by flexible compositions. *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.* **07–12–June,** 3945–3954 (2015).

11. Andriluka, M., Roth, S. & Schiele, B. Monocular 3D pose estimation and tracking by detection. *2010 IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.* 623–630 (2010). doi:10.1109/CVPR.2010.5540156

12. Yang, Y. & Ramanan, D. Articulated Human Detection with Flexible Mixtures-of-Parts. *IEEE Trans. Pattern Anal. Mach. Intell.* 1–15 (2012). doi:4E81F44D-B885-4A86-B3BB-ACD954C51B4D

13. Felzenszwalb, P. F., Girshick, R. B., Mcallester, D. & Ramanan, D. Object Detection with Discriminatively Trained Part Based Models. *IEEE Trans. Pattern Anal. Mach. Intell.* **32,** 1–20 (2009).

14. Kanade, T., Ramakrishna, V., Sheikh, Y., Kanade, T. & Sheikh, Y. Tracking Human Pose by Tracking Symmetric Parts. *Comput. Vis. Pattern Recognit. (CVPR), 2013 IEEE Conf.* 3728–3735 (2013). doi:10.1109/CVPR.2013.478

15. Wang, Y. & Mori, G. Multiple Tree Models for Occlusion and Spatial Constraints in Human Pose Estimation. *Constraints* 710–724 (2008).

16. Sapp, B. & Taskar, B. MODEC: Multimodal decomposable models for human pose estimation. *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.* 3674–3681 (2013). doi:10.1109/CVPR.2013.471

17. Komodakis, N., Paragios, N. & Tziritas, G. MRF energy minimization and beyond via dual decomposition. *IEEE Trans. Pattern Anal. Mach. Intell.* **33,** 531–552 (2011).

18. Freifeld, O., Weiss, A., Zuffi, S. & Black, M. J. Contour people: A parameterized model of 2D articulated human shape. in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* 639–646 (2010). doi:10.1109/CVPR.2010.5540154

19. Cootes, T. F., Taylor, C. J., Cooper, D. H. & Graham, J. Active Shape Models - Their Training and Application. *Comput. Vis. Image Underst.* **61,** 38–59 (1995).

20. Jiang, H. Finding human poses in videos using concurrent matching and segmentation. in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* **6492 LNCS,** 228–243 (2011).

21. Sidenbladh, H., De La Torre, F. & Black, M. J. A framework for modeling the appearance of 3D articulated figures. in *Proceedings - 4th IEEE International Conference on Automatic Face and Gesture Recognition, FG 2000* 368–375 (2000). doi:10.1109/AFGR.2000.840661

22. Anguelov, D. *et al.* SCAPE: Shape Completion and Animation of People. in *SIGGRAPH ’05 ACM SIGGRAPH 2005 Papers* 408–416 (2005). doi:10.1007/978-3-642-37484-5\_12

23. Ge, S. & Fan, G. Non-rigid articulated point set registration for human pose estimation. in *Proceedings - 2015 IEEE Winter Conference on Applications of Computer Vision, WACV 2015* 94–101 (2015). doi:10.1109/WACV.2015.20

24. Rius, I., Gonzàlez, J., Varona, J. & Xavier Roca, F. Action-specific motion prior for efficient Bayesian 3D human body tracking. *Pattern Recognit.* **42,** 2907–2921 (2009).

25. Urtasun, R., Fleet, D. J., Hertzmann, A. & Fua, P. Priors for people tracking from small training sets. in *Proceedings of the IEEE International Conference on Computer Vision* **I,** 403–410 (2005).

26. Lepetit, V. & Fua, P. Monocular Model-Based 3D Tracking of Rigid Objects: A Survey. *Found. Trends® Comput. Graph. Vis.* **1,** 1–89 (2005).

27. Ronfard, R., Schmid, C. & Triggs, B. Learning to parse pictures of people. *Proc. Eur. Conf. Comput. Vis.* **2353,** 700–714 (2002).

28. Okada, R. & Soatto, S. Relevant feature selection for human pose estimation and localization in cluttered images. in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* **5303 LNCS,** 434–445 (2008).

29. Ning, H. N. H., Xu, W. X. W., Gong, Y. G. Y. & Huang, T. Discriminative learning of visual words for 3D human pose estimation. *2008 IEEE Conf. Comput. Vis. Pattern Recognit.* 1–8 (2008). doi:10.1109/CVPR.2008.4587534

30. Tompson, J., Jain, A., LeCun, Y. & Bregler, C. Joint Training of a Convolutional Network and a Graphical Model for Human Pose Estimation. 1–9 (2014).

31. Gkioxari, G., Hariharan, B., Girshick, R. & Malik, J. R-CNNs for Pose Estimation and Action Detection. 1–8 (2014).

32. Carreira, J., Agrawal, P., Fragkiadaki, K. & Malik, J. Human pose estimation with iterative error feedback. in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* **2016–Decem,** 4733–4742 (2016).

33. Flitti, F., Bennamoun, M., Huynh, D. Q. & Owens, R. A. Probabilistic human pose recovery from 2D images. in *Proceedings - International Conference on Image Processing, ICIP* 1517–1520 (2010). doi:10.1109/ICIP.2010.5652502

34. Amit, Y. & Geman, D. Shape Quantization and Recognition with Randomized Trees. *Neural Comput.* **9,** 1545–1588 (1997).

35. Chang, J. Y. & Nam, S. W. Fast random-Forest-Based human pose estimation using a multi-Scale and cascade approach. *ETRI J.* **35,** 949–959 (2013).

36. Breiman, L. Random forests. *Mach. Learn.* **45,** 5–32 (2001).

37. Girshick, R., Shotton, J., Kohli, P., Criminisi, A. & Fitzgibbon, A. Efficient regression of general-activity human poses from depth images. in *Proceedings of the IEEE International Conference on Computer Vision* 415–422 (2011). doi:10.1109/ICCV.2011.6126270

38. Lepetit, V. & Fua, P. Keypoint recognition using randomized trees. *IEEE Trans. Pattern Anal. Mach. Intell.* **28,** 1465–1479 (2006).

39. Huang, J. Bin & Yang, M. H. Fast sparse representation with prototypes. in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* 3618–3625 (2010). doi:10.1109/CVPR.2010.5539919

40. Wright, J., Yang, A. Y., Ganesh, A., Sastry, S. S. & Ma, Y. Robust face recognition via sparse representation. *IEEE Trans. Pattern Anal. Mach. Intell.* **31,** 210–227 (2009).

41. Chen, C., Yang, Y., Nie, F. & Odobez, J. M. 3D human pose recovery from image by efficient visual feature selection. *Comput. Vis. Image Underst.* **115,** 290–299 (2011).

42. Kanaujia, A. Coupling Top-down and Bottom-up Methods for 3D Human Pose and Shape Estimation from Monocular Image Sequences. 13 (2014).

43. Agarwal, a. & Triggs, B. Monocular Human Motion Capture with a Mixture of Regressors. *2005 IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. - Work.* **3,** 72–72 (2005).

44. Sigal, L., Balan, A. O. & Black, M. J. Combined discriminative and generative articulated pose and non-rigid shape estimation. in *NIPS* **20,** 1–8 (2007).

45. Orrite-Uru??ueta, C., Del Rinc??n, J. M., Herrero-Jaraba, J. E. & Rogez, G. 2D silhouette and 3D skeletal models for human detection and tracking. in *Proceedings - International Conference on Pattern Recognition* **4,** 244–247 (2004).

46. Lu, H., Shao, X. & Xiao, Y. Pose estimation with segmentation consistency. *IEEE Trans. Image Process.* **22,** 4040–4048 (2013).

47. Eichner, M. & Ferrari, V. Better appearance models for pictorial structures. in *Procedings of the British Machine Vision Conference 2009* 3.1-3.11 (2009). doi:10.5244/C.23.3

48. Hernandez-Vela, A. *et al.* Graph cuts optimization for multi-limb human segmentation in depth maps. in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* 726–732 (2012). doi:10.1109/CVPR.2012.6247742

49. Ferrari, V., Mar??n-Jim??nez, M. & Zisserman, A. 2D human pose estimation in TV shows. in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* **5604 LNCS,** 128–147 (2009).

50. Ferrari, V., Marin-Jimenez, M. & Zisserman, A. Pose search: Retrieving people using their pose. in *2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, CVPR Workshops 2009* 1–8 (2009). doi:10.1109/CVPRW.2009.5206495

51. Ladicky, L., Torr, P. H. S. & Zisserman, A. Human Pose Estimation Using a Joint Pixel-wise and Part-wise Formulation. *Comput. Vis. Pattern Recognit. (CVPR), 2013 IEEE Conf.* 3578–3585 (2013). doi:10.1109/CVPR.2013.459

52. Mori, G. Guiding model search using segmentation. in *Proceedings of the IEEE International Conference on Computer Vision* **II,** 1417–1423 (2005).

53. Nakariyakul, S. A comparative study of suboptimal branch and bound algorithms. *Inf. Sci. (Ny).* **278,** 545–554 (2014).

54. Tian, T. P. & Sclaroff, S. Fast globally optimal 2D human detection with loopy graph models. in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* 81–88 (2010). doi:10.1109/CVPR.2010.5540227

55. Jiang, H. Finding Human Poses in Videos. 228–243 (2011).

56. Rogez, G., Rihan, J., Ramalingam, S., Orrite, C. & Torr, P. H. S. Randomized trees for human pose detection. in *26th IEEE Conference on Computer Vision and Pattern Recognition, CVPR* (2008). doi:10.1109/CVPR.2008.4587617

57. Forsyth, D. Object detection with discriminatively trained part-based models. *Computer (Long. Beach. Calif).* **47,** 6–7 (2014).

58. Rothrock, B., Park, S. & Zhu, S. C. Integrating grammar and segmentation for human pose estimation. in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* 3214–3221 (2013). doi:10.1109/CVPR.2013.413

59. Torres Garcia, F. & Kropatsch, W. Top-down 3D Tracking and Pose Estimation of a Die Using Check-points. in *Proceedings of the 18th Computer Vision Winter Workshop* 102–109 (2013).

60. Bray, M., Kohli, P. & Torr, P. PoseCut: simultaneous segmentation and 3D pose estimation of humans using dynamic graph-cuts. *Network* **1,** 642–655 (2006).

61. Ionescu, C., Li, F. & Sminchisescu, C. Latent structured models for human pose estimation. in *Computer Vision (ICCV), 2011 IEEE International Conference on* 2220–2227 (2011). doi:10.1109/ICCV.2011.6126500

62. Kolmogorov, V. & Zabih, R. What Energy Functions Can Be Minimized via Graph Cuts? *IEEE Trans. Pattern Anal. Mach. Intell.* **26,** 147–159 (2004).