南華大學
UNIVERSITY OF SOUTH CHINA

# 毕业论文文献综述

题　　目　基于最优传输理论图像匹配算法文献综述

学院名称　数理学院

指导教师　高有

职　　称　讲师

班　　级　信计 1802 班

学　　号　20184390213

学生姓名　朱柳承

2022 年 1 月 30 日

# 基于最优传输理论图像匹配算法文献综述

**摘要**：图像匹配是虚拟图像重建的一个关键步骤，是视觉识别的一个核心过程，是图像检索的中心操作模块。视觉算法在过去几十年来飞速发展，基于图像匹配的应用更是层出不穷，从卫星遥感图像配准到纳米级零件配准，在科学研究、社会需求、工业制造等各个方面，图像匹配这一基本问题无处不在。而对于图像匹配其定义一般被描述如下：同一摄影项目的图像如医学图像，可以在遮挡、多姿态等条件下从任何的光照强度和频率以各种角度拍摄，这些同一项目的图像之间存在内容、结构、特征、色彩及纹理等对应关系，而图像匹配则致力于这些关系中的一致性和相似性分析。

尤其随着计算机运算能力的提升、图像处理加速芯片的发展，图像匹配算法数量越来越多，其类型也越来越丰富。随之产生了两个问题：新方法与旧方法的取舍；如何在现有理论指导下设计更适用、准确、鲁棒的高性能匹配算法。为了回答这两个问题，我们有必要系统的回顾和评估过去和现在的图像匹配算法。我在检索相关的文献时，发现了一个奇怪的现象：如今大热的并且在理论上对于图像匹配算法的发展具有巨大的推进潜力的最优传输理论方面的内容常常被一些重要的文献综述 [2] 所忽略。因此，本文主要讨论基于最优传输理论的图像匹配算法。

本文首先回顾了遵循着从手工设计特征到深度学习的图像匹配算法的发展路线，并简要分析了其中各个阶段的算法的特点。然后，详细地介绍最优传输理论在视觉算法上的进展，其中涉及计算最优传输映射的方法、计算 Wasserstein 距离的方法的概述；接着，列举几个基于最优传输理论的图像匹配应用，并且与经典的解决方案相比较，以此了解 OT 在视觉算法中的意义。最后，我们总结了图像匹配技术的现状，并对未来的工作进行了富有洞察力的讨论和展望。本调查可作为（但不限于）图像匹配及相关领域的研究人员和工程师参考。

**关键词：**图像匹配，共形映射，最优传输理论，最佳质量传输映射，Wasserstein 距离，曲面配准，手工设计特征，深度学习

# Graduation Thesis Literature Review

## Graduation Thesis Topic: Literature review of image matching algorithms based on optimal transmission theory

Student name: Zhu Liucheng Student number: 20184390213

Tutor name: Gaoyou Professional qualifications: Lecturer

**Abstract** : Image matching is a key step in virtual image reconstruction, a core process of visual recognition, and a central operation module of image retrieval. With the rapid development of visual algorithms in the past decades, image matching-based applications have emerged one after another. From satellite remote sensing image registration to nano-scale parts registration, the basic problem of image matching is ubiquitous in scientific research, social needs, industrial manufacturing and so on. The definition of image matching is generally described as follows: images of the same photographic project, such as medical images, can be taken from any angle of light intensity and frequency under occlusion, multi-posture, etc. There are corresponding relationships among the images of the same project, such as content, structure, feature, color and texture, while image matching focuses on consistency and similarity analysis in these relationships.

Especially with the improvement of computer computing ability and the development of image processing acceleration chip, there are more and more image matching algorithms and their types are more and more abundant. Two problems arise: the choice between the new method and the old one; How to design a more applicable, accurate and robust high performance matching algorithm under the guidance of the existing theory. In order to answer these two questions, it is necessary to systematically review and evaluate past and present image matching algorithms. When I retrieve the relevant literature, I found a strange phenomenon: the hot and theoretically promising aspects of optimal transmission theory for the development of image matching algorithms are often ignored by some important literature reviews[2]. Therefore, this paper mainly discusses image matching algorithms based on optimal transmission theory.

This paper first reviews the development of image matching algorithms that follow the path from manual design features to in-depth learning, and briefly analyses the characteristics of the algorithms at each stage. Then, the progress of optimal transmission theory in visual algorithms is introduced in detail, including an overview of the methods for calculating optimal transmission mappings and Wasserstein distances. Next, several image matching applications based on optimal transmission theory are listed and compared with classical solutions to understand the significance of OT in visual algorithms. Finally, we summarize the current status of image matching technology, and make insightful discussions and prospects for future work. This survey can be used as a reference for (but not limited to) researchers and engineers in image matching and related fields.

**Keywords:** Image matching, Conformal mapping, Optimal transmission theory, Optimal mass transport mapping, Wasserstein distance, Surface registration, Handcrafted feature, Deep learing

# 1 引言

元宇宙概念在近年来吸引了大量企业与研究者的关注，主要是因为其在技术上，特别是与虚拟图像、计算机辅助技术领域的三大突破有关。其一是英伟达发布了世界首款实时光线追踪 GPU，我们知道高质量的 3D 渲染的核心算法是基于几何光学的光线追踪法。二十年前，该算法只能在昂贵的 Sun 或者 SGI 工作站上运算。依随岁月的流逝，越来越多的物理定则被加入到算法流程之中，渲染效果愈发逼真。几乎所有的电影特效都是基于光学追踪法，一部电影往往需要数千台 Linux 服务器计算数年。长久以来，大家都将实时光线追踪计算作为一个梦想。终于，英伟达的 GPU 技术积累到达了这个临界点。

其二便是 Epic Game 发布的虚幻引擎五，它具备两大全新核心技术：Nanite 虚拟微多边形几何技术和 Lumen 动态全局光照技术。Nanite 虚拟几何技术的出现意味着由数以亿计的多边形组成的影视级艺术作品可以被直接导入虚幻引擎，Nanite 几何体可以被实时流送和缩放，因此无需再考虑多边形数量预算、多边形内存预算或绘制次数预算了；也不用再将细节烘焙到法线贴图或手动编辑细节层次（LOD），这必定是图形学领域革命性的飞跃。

其三便是 AI 的 GAN model，对抗生成网络（Generative Adersarial Network GAN）获得了爆炸式的增长，其应用范围几乎涵盖了图像处理和机器视觉的绝大多数领域。其精妙独到的构思，令人拍案叫绝；其绚烂逼真的效果，令众生颠倒。一时间对抗生成网络引发了澎湃汹涌的技术风潮，纳什均衡的概念风靡了整个人工智能领域。GAN 的核心思想是构造两个深度神经网络：判别器 D 和生成器 G，用户为 GAN 提供一些真实货币作为训练样本，生成器 G 生成假币来欺骗判别器 D，判别器 D 判断一张货币是否来自真实样本还是 G 生成的伪币；判别器和生成器交替训练，能力在博弈中同步提高，最后达到平衡点的时候判别器无法区分样本的真伪，生成器的伪造功能炉火纯青，生成的货币几可乱真。这种阴阳互补，相克相生的设计理念为 GAN 的学说增添了魅力。

## 1.1 全局准则

图像的一般描述方法有纹理、统计、基于模型以及基空间方法。纹理作为一个关键度量，是图像处理中的重要主题，他通常分为结构方法和统计方法，结构方法寻找边缘和形状等特征，而统计方法关心的是像素值的关系和统计矩。

Fourier 空间等基空间方法也可以用于特征描述。在 20 世纪 60 80 年代，人们为了在高分辨率的彩色图像上做一些匹配等任务，往往只有在内存足够的情况下才能进行。这一时期主要是整体目标方法，它用特征度量来描述几乎整个目标、较大的区域或图像。大型目标的模式匹配采用 FFT 谱方法和其他方法，识别方法包括目标、形状以及纹理等度量，并使用简单的集合元素进行目标组合。NTSC 制、PAL 制和 SECAM 制等低分辨率图像比较常见，而且主要是灰度图像。

基于上述的全局特征与一些几何方法，通过计算两幅图像的特征向量之间的欧氏距离来进行匹配。这种方法本质上对光照变化具有鲁棒性，但有一个巨大的缺点：即使使用最先进的算法，标记点的准确配准也很复杂。在 [2] 中进行了一些关于几何人脸匹配的工作. 使用了 22 维特征向量，并且对大型数据集的实验表明，仅几何特征可能无法携带足够的信息来进行人脸匹配。简而言之这类图像匹配算法缺乏泛化能力。于是人们在分析图像空间时考虑了降维方法，认为高维的图像空间如 $100 \times 100$ 的图像即 10000 维的图像空间中，只有部分像素是我们关心的，因此主成分分析法很自然地引入。在图像匹配过程中，该分析识别具有最大方差的轴，即确定一组正交基从而得到描述图像的主成分向量，接着把两幅图像的主成分之间的距离作为相似性度量进行图像匹配操作。

然而，PCA 方法从重建的角度来看，这种转换是最佳的，但它没有考虑任何类标签。比方说对一个人脸进行差异化分类，可分为脸间差异和脸内差异。脸内差异表示同一个人脸的各种可能变形。脸间差异表示不同人的本质差异。对于同一个人，不同表情会使匹配效果不稳定，因此一般 PCA 方法会使用同一项目的各种姿势下的平均图像作为匹配依据，但是当一个方差是由外部来源产生的情况下，具有最大方差的轴不一定包含任何判别信息，此时分类变得不可能。对于这一类情况，人们进一步提出了将具有线性判别分析的特定类别投影应用于人脸匹配 [3]，其基本思想是最小化类内的方差，同时最大化类间的方差。

上述的图像匹配思想都是基于全局特征进行的，这有一个不足之处便是当图像的热点区域若出现遮挡或缺损等情况，上述算法则会不稳定。因此，人们进一步提出了各种基于局部特征提取的方法。

## 1.2 局部特征准则

20 世纪 90 年代初期（部分目标方法），人们越来越多地使用局部特征和兴趣点来描述图像中较小的目标、目标的部件和图像区域，例如 Shi 和 Tomasi[149] 改进了 Harris 检测器，kitchen 和 Rosenfeld[200] 提出了灰度角点检测方法，Khotanzad 和 Hong[268] 与 1990 年提出使用 Zernike 多项式来计算多边形形状的图像矩等。20 世纪 90 年代中期（局部特征方法）：特征描述子从每个特征周围的窗口上添加更多细节，通过特征搜索和匹配来进行目标识别。先是搜索特征集并使用更复杂的分类器来匹配描述子。描述子包括梯度、边缘和颜色。20 世纪 90 年代后期，开发了各种具有局部不变性的特征描述子，这些描述子对尺度、亮度、旋转和仿射变换等具有不变性。Schmid 和 Mohr[340] 详细介绍了局部特征描述方法。特征就像字母，是拼写出复杂的特征描述子或向量的基础，这些向量将用于特征匹配。

21 世纪初期 Lowe 提出的 SIFT[153] 算法和 Bay 等人提出的 SURF[152] 算法都采取了不同的方式来使用 HAAR 特征而不仅是特度特征。2010 年以后：多模态特征度量融合，这一时期人们更多地使用深度传感器信息和深度图来分割图像、描述特征，比如 Rusu 和 Bradski 等人创建了 VOXEL 度量 [380] 和 2D 纹理度量在 3D 空间中的表示。人们开发出更快、更好的二值模式特征描述子，这种描述子使用汉明距离进行快速匹配，如由 Alahi 等人提出的 FREAK[122]、由 Rublee 等人提出的 ORB[112]。多模态和多变量描述子由图像特征和其他传感器（比如加速度计传感器、位置传感器等）信息构成。

只描述图像的局部区域，提取的特征对部分遮挡、光照和小样本量更鲁棒。用于局部特征提取的算法有 Gabor Wavelets ([4])、离散余弦变换 ([5]) 和局部二进制模式 ([6])。在应用局部特征提取时，什么是保留空间信息的最佳方法仍然是一个开放的研究问题，因为空间信息是潜在的有用信息。换句话说选择一些具有鲁棒性的特征算子仍然是困难的问题。人们对此的解决方案，是采用机器学习的思想来选取更好的特征，

## 1.3 特征学习准则

21 世纪初期，这一时期采用有良好形势的描述子将场景和目标建模为特征组件或模式集合；为了进行特征匹配，需度量特征之间的空间关系；新的复杂的

分类和匹配方法会采用 Boosting 及相关方法，这种方法结合强弱特征来进行更优效的识别。Viola 和 Jones 方法 [486] 使用 HAAR 特征和基于 Bossting 的学习方法来进行分类，从而加快了匹配速度。21 世纪头十年中期（较细粒度的特征和度量组合方法），Czuka 等人 [226] 提出描述场景和目标的各种特征与度量的组合，而 Sivic 采用关键点方法来描述场景。此外人们还开发了特征学习和稀疏特征码本方法，以减少模式空间，加快搜索速度、提高准确性。

特征学习方法会会创建一组平均的、被压缩（或稀疏）的分层特征集，这就是训练集中的主要特征。机器学习过程可以表述为如下 3 种方式：(1) 特征提取：可用局部特征描述子或通过深度神经网络来学习特征。(2) 特征编码：可保留所有特征集或仅保留一个稀疏特征集。(3) 分类器的设计和训练。

特征学习架构包括以下两大类：(1) 统计学习方法。这类方法包括在特征描述子中广泛使用的方法、学习方法、稀疏编码和统计分类器。(2) 神经网络方法。它是受神经生物学概念的启发而建立的，比如局部感受野与人工神经元的连接、深度特征层次。

其中基于 Haar 特征的 Adaboost 算法取得目前在室内外环境中位姿估计任务中最稳定的结果。该类算法展示了基于注意力的图神经网络对局部特征匹配的强大功能。SuperGlue 的框架使用两种注意力 ([7])：（i）自我注意力，可以增强局部描述符的接受力；（ii）交叉注意力，可以实现跨图像交流，并受到人类来回观察方式的启发进行匹配图像。文中方法通过解决最优传输问题，优雅地处理了特征分配问题以及遮挡点。实验表明，SuperGlue 与现有方法相比有了显着改进，可以对室内和室外的图像进行高精度的相对姿势估计。此外，SuperGlue 可以实时运行，并且可以同时使用经典和深度神经网络去学习特征。如今，深度学习在图像匹配中应用地更为广泛，但是它的不足之处也是显然的，模型的可解释性差、模型坍塌等等。

## 1.4 总结

深度学习得到的特征为什么效果会好呢? 其原因有：(1) 特征的绝对数量；(2) 特征的层次性，即特征能表示低级概念、中级概念和高级概念。这表明用层次化方法来创建局部特征集（比如 SIFT 和 FREAK）有可能得到与卷积神经网络（采用简单相关性模板特征）相当（甚至可以超过）的性能。

# 2 基于最佳质量运输理论的匹配算法

最优传输理论起源于两百多年前，由 Monge 提出的经典问题——确定以最小运输成本将一堆沙子从这个地方移动到另一个地方的最佳方式 [9]。Kantorovich [31] 证明了基于线性规划的最优运输计划的存在性和唯一性。Monge-Kantorovich 优化已被用于从物理学、计量经济学到计算机科学的众多领域，包括数据压缩和图像处理 [41]。最近十年，研究人员已经意识到，如果可以降低其高计算成本 [16]、[54] ，最优传输可以为图像处理提供强大的工具。. 但是，它有一个基本缺点，即变量的数量是（ķ2)，这对于计算机视觉和医学成像应用来说是不可接受的，因为高分辨率 3D 表面通常包含多达数十万个顶点。另一种 Monge-Brenier 优化方案可以显着减少要优化的变量数量。在 1980 年代后期，Brenier [11] 为一类特殊的最优运输问题开发了一种不同的方法，其中成本函数是二次距离。Brenier 的理论表明，最优传输图是特殊凸函数的梯度图。假设目标域离散化为 n 样本，Monge-Brenier 的方法减少了未知变量（n2) 到 o ( n )，大大降低了计算成本，提高了效率。

目前，最优传输理论之所以在计算机视觉领域开始发光发热，主要是因为该理论的计算机算法得到了极大的进展，求解一个 OT 问题不再是困难的问题。另一个方面是，最优传输理论主要研究的是两个分布之间的最优传输映射，而图像在计算机科学界公认的一个流形分布定则的解释下，可以认为是嵌在高维图像空间下的概率分布，而两幅或多幅图像之间的匹配，其实就是计算两个分布之间的运输成本，换句话说即是计算两个分布之间的最优传输映射，并根据映射函数的运算结果去分类。当然，这里有几个不可忽视的问题，一、经典的 OT 映射，需要两个分布之间的映射是质量守恒的，即运输过程中没有损失质量，用数学语言描述映射是保测度的。二、经典的 Wasserstein 距离计算的是同一图像空间下的概率分布，如果是不同类的图像空间，该距离度量将会失效。三、尽管现如今有许多加速或简化 OT 映射计算的算法，但是速度较快的如 Sinkhon 它计算的其实是一个正则化的简化版的伪 OT 映射，即它计算结果只能作为最优结果的近似。而 OT 映射的数值解法一部分是运算复杂度高，一部分是理论晦涩难懂算法实现困难。因此，研究人员一般会把 OT 映射的计算按照特定的任务使用相应的定理，并采用一些如牛顿法、最速下降算法去求取推导出的一些公式，以此逼近 OT 映射。总而言之，OT 映射计算算法的选择也是一个需要关注的问题。鉴

于上述的各种情况，目前，OT 理论在视觉任务中的应用与发展一般分为四个方面：预处理、后处理、Wasserstein 距离、解释 GAN 及相似模型并设计基于 OT 的匹配算法。

## 2.1 Wasserstein 距离

假设 $(M, g)$ 是一个黎曼流形，其黎曼度量为 $g$。

**定义 2.1.** 设 $\mathcal{P}_p(M)$ 表示 $M$ 上具有有限 $pth$ 矩的所有概率测度 $\mu$ 的空间，其中 $p \geq 1$。假设存在某一点 $x_0 \in M$，有 $\int_M d(x, x_0)^P \mathrm{d}\mu(x) < +\infty$，其中 $d$ 是 $g$ 的测地距离。

给定 $\mathcal{P}_p$ 中的两个概率测度 $\mu$ 和 $\nu$，它们之间的 Wasserstein 距离被顶为最佳质量运输映射 $T : M \to M$ 的运输成本，

$$W_p(\mu, \nu) := \inf_{T_{\#}\mu=\nu} \left( \int_M d(x, T(x))^p \mathrm{d}\mu(x) \right)^{\frac{1}{p}}. \tag{1}$$

下面的定理对当前的工作起着基础性作用。

**定理 2.1.** *Wasserstein 距离 $W_p$ 是 Wasserstein 空间 $\mathcal{P}_p(M)$ 的黎曼度量，详细证明间参考文献 [55]。*

Wasserstein 距离不仅给出了两个分布之间的距离，而且能够告诉我们它们具体如何不一样，即如何从一个分布转化为另一个分布，靠的就是联合分布 $(x, T(x))$。Wasserstein 距离之所以很难计算，一个重要原因是两个分布的维度一般成百上千，如果用传统的线性规划算法求解该问题，计算复杂度是难以承受的。这成为了一个限制其应用的难点目前能用显式计算出来的只有两种情况一种是 1 维即 $d = 1, p = 1$，另一种是高斯分布。当然，在精确解上有困难，我们可以考虑求一个近似解，其中利用计算几何中的工具，可以有效地解决最佳运输的具体实例。例如，从连续到离散点状措施的运输成本可以通过多尺度算法计算 [7]，或通过欧几里得空间上的牛顿迭代 [de Goes et al. 2012; Zhao et al. 2013] 最近这种基于牛顿法的计算方法被扩展到离散曲面 [de Goes et, 2014]. 点云和线段之间的传输距离也在二维中基于平面的三角平铺和贪婪点到线段聚类来近似 [de Goes ,2011]。另一项工作提出了一个具有额外时间变量的最优运输的动态公式，对于距离成本的平方，Benamou 和 Brenier[2000] 通过将一个分部在时间上平流

到另一个分部的成本最小化来计算运输距离。对于非平方距离代价，Solomon 等人 [2014a] 将运输映射求解为向量场的流，其散度与输入密度的差值相匹配。

其他方法使用最优运输从多个密度聚集和平均信息。例如重心计算 [Agueh 和 carlier 2011]，图上的密度传播 [Solomon 等人 2014b]，以及 “软” 对应映射的计算 [Solomon 等人 2012]。这些问题通常通过一个多边际线性程序来解决 [Agueh 和 carlier2011;Kim 和 pass2013]，这对于大规模域是不可行的。一种变通的方法是使用带有亚梯度方向的 L-BFGS 来处理线性规划的对偶 [Carlier 等人 2014] 但这种策略存在条件反射差和噪声结果的问题。

正则化为运输问题的近似求解提供了一种可能，长期以往，内点法一直使用障碍函数将线性规划转化为严格的凸问题，而熵正则化在最优运输的特殊情况下提供了 [Cuturi 2013] 中所述的几个关键优势在熵正则化的情况下，使用迭代比例拟合 (IPFP) 或 Sinkhorn-Knopp 算法来解决最优运输问题 [Deming and Stephan 1840; Sinkhorn 1976]，它可以在并行的 GPU 架构中实现，并用于计算如上千个分布的重心 [Cuturi 和 Doucet 2014]。此外，有一部分工作 [Convolution...] 利用迭代标度方法解决熵正则化输运和相关问题的效率 [Cu-turi 2013;Benamou 等 2015]。通过在连续语言中提出正则化传输，并将这些算法的效率与曲面和图像等领域的离散化相结合，这种变化不仅仅是符号上的，而是带来了更快的迭代图像上的高斯核与表面的热核相连接；这些核可以在不预先计算成对距离矩阵的情况下进行计算。

## 2.2 OT 方法

OT 理论同样可以在图像匹配的前置处理和后置处理发挥作用。高精度点云配准任务 [Accurate Point Cloud Registration with Robust Optimal Transport]，我们可以把这个复杂任务描述为两个问题：(1) 如何寻找 source 和 target 的对应关系；(2) 如何利用这个对应关系求解特定的变换。对于 (1) 隐性地利用到了相似性的概念，直观来说就是特征匹配：用 source 的点去匹配 target 中的点，并计算他们的特征相似度。这里可以引入 Entropic 和 unbalanced 形式下的 OT，文献中称为 RobOT，并引入 Weighted RobOT Matching，具体是通过计算 target set 在 $\pi_i j$ 加权下的 barycenter，得到加权的目标位置。到这里为止，我们简单的描述了如何利用 RobOT 直接求解配准问题。没有引入其他的 regularization 项，但这种解决方

案并不是完美的。如下例子目标是把彩色月亮匹配到蓝色月亮上去，我们简单的采用 xyz 作为每一个点的特征。虽然形状匹配的很好，但很明显它的拓扑结构乱了。这是因为 RobOT 本身并没有 regularize transformation, 这个在有较大 rotation 的情况下非常明显。因为我们还加入了非 deep learning based 前置预处理模块和后置 finetune 模块。这两个模块都是基于 RobOT 所以速度在毫秒级，并不影响整体的速度。总体而言，我们用了一个预处理模块（optimiatzion-based），deep non-parametric 模块（支持多种 deformation model，deep-learning based), 后处理模块（optimization-baed) 的三明治结构。我们称这个方法为 Deep-RobOT(D-RobOT)

至于为什么这么设计呢。三个模块对应三个原因 1）为什么加入预处理，一般而言大部分 non-parametric 模型都是针对与局部形变设计的 regularization，需要预先移除全局 translation，rotation, shear 和 scale。2）为什么采用 non-parametric，局部形变的处理需要 non-parametric 模型，另外通过任务类型可以选择相应的模型，像场景流这类型小形变任务，control point based spline model 能求解平滑的流场，对比 full resolution 的位移场，spline 的结果一般更平滑，另外 control point 的引入可以大大降低计算内存消耗。3）为什么采用后处理，这是因为不要过分的依赖深度模型，很多时候模型可能并不会给出完美的结果，这时候后处理可以修正模型误差。

除了上述例子之外，把 OT 方法用在匹配任务的前置或后置处理的例子还有很多。其中一个最为经典的例子便是 [4]，利用 OMT-Map 实现一个区域保持映射，将源和目标映射到一个单位平面圆盘上，而不会产生大面积失真。然后将表面配准问题转化为两个平面圆盘之间的地标匹配 T-Map，使映射在整个域上具有一致的形象失真，同时使最大形象失真最小。

## 2.3 OT 匹配算法

当前 OT 理论主要从两个方面启发图像匹配算法的设计，一是求取图像之间 OT 映射，二是 OT 理论在深度学习中的阐释。对于求取图像之间的 OT 映射，最为引人注意的便是医学图像上的配准以及 3D 点云的匹配。如 [5] 利用保角映射将具有圆盘拓扑的度量曲面映射到单位平面圆盘上，利用所产生的面积畸变获取其概率测度，通过计算具有两个概率测度的两个表面之间的唯一的最优质量运输映射（用牛顿法进行能量优化），我们可以得到定义两个表面之间的 Wasserstein

距离的最优运输成本，这种 Wasserstein 距离可以在本质上测量基于曲面形状之间的差异，从而可以用于形状分类。再如 [9] 通过最小化两个域间联合分布的 Wasserstein 距离来解决上述挑战．提出一个定理将难以求解的 Wasserstein 距离原问题转化为一个简单的优化问题，并设计了一个联合 Wasserstein 自编码器模型（JWAE）来求解该问题．然后，本文将 JWAE 成功应用在无监督图像翻译和跨域视频合成任务中，并生成高质量的图像和连贯的视频。

而对于 OT 理论在深度学习中的阐释，根据流形分布定则和流形嵌入定理 [....] 可以把数据集描述为数据流形，而对于图像数据集，它所表示的数据流形 $\sum$ 嵌入在图像空间 $\mathbb{R}^n$ 中。而数据集可以被抽象成一个数据流形上的概率分布 $\mu$，编码映射 $\varphi_i : U_i \to \mathcal{Z}$ 将数据流形上的一个领域 $U_i$ 映射到隐空间 $\mathcal{Z}$ 上。换句话说深度学习中的编码映射和解码映射本质上是将陆行嵌入到不同维数的欧氏空间中，如果流形的嵌入具有扭结结构，通过嵌入到不同维数的欧式空间可以解除扭结结构；如果初始流形嵌入的维数过高，通过改变嵌入空间而实现逐步降维，直至隐空间。那么这套说法解释了为什么可以把图像描述成一个概率分布，即深度学习的学习对象。那么深度学习为什么可以得到能够描述这些流形分布的隐空间的呢？一种说法是万有逼近定理 [Weierstrass]，即深度学习是复合简单函数来逼近任意的连续函数和连续映射的基于如下定理

**定理 2.2.** 假设 $f$ 是一个多元连续函数，那么 $f$ 可以被写成单元连续函数的有限复合，

$$f(x_1, x_2, \cdots, x_n) = \sum_{q=0}^{2n} \Phi_q \left( \sum_{q=1}^{n} \phi_{p,q}(x_p) \right)$$

这里 $\phi, \Phi$ 分别称为内、外函数

我们有多种方式用深度神经网络来构造内、外函数，例如用 Sigmoid、ReLU 激活函数来表示内函数。由此可知，深度学习训练出隐空间的过程即编码过程其实是一个把原分布映射到隐空间的过程，然后再通过解码器映射到图像空间中，这里研究人们把一些经典的深度学习模型如 GAN model，使用 OT 的思想修改它的某个步骤，从而得到一些有效的图像匹配算法。

# 3 OT 匹配算法的应用

图像匹配是计算机视觉中的一个基本问题，被认为是广泛应用中的先决条件。其中具有代表性的主要有 1. SFM(Structure-from-motion)，从一系列图像中恢复静止场景的 3D 结构。2. SLAM(Simultaneous Localization and Mapping)，同时定位和映射。3. Visual Homing，旨在仅基于信息将机器人从任意起始位置导航到目标或原始位置。4. 图像配准与变形，在 [1] 中详细介绍了 OT 在实际成像问题中，所做的图像配准和变形的步骤，其中的一个关键之处，便是在给出了一个精确 Monge-Kantorovich 问题的公式然后通过寻找保质量映射的极坐标分解的等价问题来找到最优映射 [Gangbo 1994 ; makeken, 2001]，可通过自然梯度下降算法来实现，然后在函数中加入一个比较项来惩罚强度的变化。这样一来，便能适用于图像变形。又如 [8] 中通过改进仰卧位到俯卧位结肠的配准框架，实现了基于最优质量运输的可视化；通过在新的目标测量密度中引入高斯曲率，设计了一个基于 OMT-Map 的矩形域目标测量密度，使得息肉可以被放大，从而更好地显示息肉，并使用颜色去编码距离的变化。以及在 [5] 中，建议使用最佳质量传输映射进行形状的匹配和比较。因为基于 Monge-Brenier 的 OT 方法它们的归一化保形映射和 OT 映射是唯一的，没有重新参数化的歧义优于弹性形状度量方法。还可以定制唯一的 OT 映射穷尽整个图像的微分同胚组。此外有更好的适用性。

5. 图像合成，在 [9] 中，通过两个不同域之间进行线性插值和翻译将联合 Wasserstein 自编码器应用在跨视频合成任务中。在 [6] 介绍了如何使用 OT 映射匹配和合成简化图像。该算法将含有噪声和离群点的缺陷点集作为输入，其中输入点集被认为是 Dirac 测度的和，而单纯复形被认为是 0-和 1-单形上的一致测度的和。通过对输入点集的 Delaunay 三角剖分进行贪婪抽取，设计了一种由细到粗的方案来构造所得到的单纯复形。

6. 图像检索、对象识别和跟踪，OT 的匹配算法也在某些成像应用的背景下进行了研究，特别是基于内容的图像检索 [Rubner1999; Rubner 1998 Levina Bickel2001]。在这项工作者，图像的像素根据其颜色位置和空间位置被分成几个 bins（或称为“签名”）。通过计算两幅图像之间的 Wasserstein 距离进行图像检索。在 [3] 中提出了一种基于 Monge-Brenier 的 OMT 理论的 Wasserstein 距离方法用于癫痫患者大脑海马的形状分类。该方法首先利用保角映射将具有圆盘拓扑结构的度量曲面映射到单位平面圆盘上，然后通过诱发的面积畸变获得概率测度。

通过计算具有两个概率测度的两个表面之间唯一的 OT 映射，可以得到定义两个表面之间的 Wasserstein 距离的 OT 成本。该 Wasserstein 距离本质上可以衡量基于形状的表面之间的差异，因此可以用于形状分类。

# 4 可选的加速方案和评估方法

## 4.1 可选的加速方案

有很多常用的加速方法可应用到计算机视觉流程中，包括监控内存管理、使用线程进行粗粒度并行、使用 SIMD 和 SIMT 方法的数据级并行、多核并行、高级的 CPU 和 GPU 汇编语言指令，硬件加速器等。有两种基本的加速方法：

1. 针对数据的加速方法；

2. 针对算法的加速方法。

计算设备的优化算法也称为流处理，是设计时通常会考虑的。但优化数据流和数据驻留可得到更好的结果，例如，在计算资源来回传递数据和格式化数据不是好的做法，这样的数据复制和格式转换会花很多时间和功耗。在慢的系统内存中复制数据比通过计算单元的快速寄存器来访问数据要慢得多。这需要基于内存速度来考虑内存架构的层次，考虑计算机视觉中图像的密集型特征，最好是找到跟踪数据的方法，并将数据尽可能长时间驻留在快速寄存器和高速缓冲区中。（可例 sinkhorn）

## 4.2 可选的评估方法

对于医学图像等匹配算法评估方法，

1. 距离误差 2. 曲率差 3. 区域失真评价 4. 不同目标措施的比较 5. 目标措施效果的可视化 6. 视觉配准评估 7. 消融实验

而对于视频流上的图像匹配算法评估方法 ( 1) Inception Score(IS)[38] 广泛用于生成模型中.IS 通过利用 Inception-V3 模型 [3 的类别预测信息来评估生成样本的质量和多样性. (2) Frechet Inception Distance(FID)[40] 也是一个广泛使用在生成模型上的指标.FID 可以评估生成图像的质量, 因为它能捕获生成样本与真实样本的相似性, 并与人类判断相关联. (3)Video variant of FID(FID4Video)[41] 评估视频的质量和连贯性. 本文使用一个预训练的视频识别模型 I3D1] 对视频序列提取

特征. 然后, 对这些特征计算 FID4 Video. 一般而言, 对于 IS 指标, 值越大代表着翻译的图像质量越好F对于 FID 和 FID4 Video 这两种指标, 值越小意味着翻译的图像或视频的质量越好.

# 5 总结与展望

对于可预见的未来，设计图像检索工具的主要限制是我们对视觉的理解十分有限。尽管理解不充分，我们也能构造有用的工具，就如 IBM 的已经出现在大量的市场广告中的图像搜素产品 QBIC，以及看上去生意兴隆的 Virage 公司的图像搜索引擎等等，但是我们仍然很难去评价怎样算成功。表示图像的方式粗略地分有三种：在像素级，人们对具体的箱数值感兴趣；在组合级，人们关心图像的整体外观；或是在对象语义级，人们关注图像所描述的事务。

# 参考文献


[1] Haker S, Zhu L, Tannenbaum A, et al. Optimal mass transport for registration and warping [J]. International Journal of computer vision, 2004, 60(3): 225-240.

[2] Ma J, Jiang X, Fan A, et al. Image matching from handcrafted to deep features: A survey [J]. International Journal of Computer Vision, 2021, 129(1): 23-79.

[3] Ma M, Lei N, Su K, et al. Surface-based shape classification using wasserstein distance [J]. Geometry, Imaging and Computing, 2015, 2(4): 237-255.

[4] Ma M, Lei N, Chen W, et al. Robust surface registration using optimal mass transport and teichmüller mapping [J]. Graphical models, 2017, 90: 13-23.

[5] Su Z, Wang Y, Shi R, et al. Optimal mass transport for shape matching and comparison [J]. IEEE transactions on pattern analysis and machine intelligence, 2015, 37(11): 2246-2259.

[6] De Goes F, Cohen-Steiner D, Alliez P, et al. An optimal transport approach to robust reconstruction and simplification of 2d shapes [C]//Computer Graphics Forum: volume 30. Wiley Online Library, 2011: 1593-1602.

[7] Mérigot Q. A multiscale approach to optimal transport [C]//Computer Graphics Forum: volume 30. Wiley Online Library, 2011: 1583-1592.

[8] Ma M, Marino J, Nadeem S, et al. Supine to prone colon registration and visualization based on optimal mass transport [J]. Graphical Models, 2019, 104: 101031.

[9] 曹杰彰, 莫朗元, 杜卿, 等. 基于最优传输理论的联合分布匹配方法及应用 [J]. 计算机学报, 2021, 44(06): 1233-1245.

[10] Villani C. Optimal transport: old and new: volume 338 [M]. Springer, 2009.

[11] Peyré G, Cuturi M, et al. Computational optimal transport [J]. Center for Research in Economics and Statistics Working Papers, 2017(2017-86).