基于几何保持图像到图像平移的无监督多模态图像配准

         Moab Arar Yiftach Ginger Dov Danon Amit H.Bermano Daniel Cohen或

特拉维夫大学

摘要

*许多应用，如自动驾驶，严重依赖于多模态数据，其中需要模态之间的空间对齐。大多数多模态配准方法都难以使用流行的跨模态相似性度量来计算图像之间的空间对应关系。在这项工作中，我们通过在两种输入模式上训练图像到图像的翻译网络，绕过了开发跨模式相似性度量的困难。这种学习过的翻译允许使用简单可靠的单模态度量来训练注册网络。我们使用两个网络——空间变换网络和平移网络——进行多模态配准。我们证明，通过鼓励我们的翻译网络保持几何结构，我们能够训练出一个精确的空间转换网络。与最先进的多模态方法相比，我们提出的方法是无监督的，不需要成对的对齐模态进行训练，并且可以适应任何成对的模态。我们在商业数据集上对我们的方法进行了定量和定性评估，表明它在几种模式上表现良好，并实现了精确的校准。*

# 1.  介绍

|  |
| --- |
|  |
|  |  |

从医学成像和遥感等经典传感器到自动驾驶等新兴任务，使用不同传感器获取场景是各个学科的常见做法。多模态传感器允许收集广泛的物理特性，从而产生更丰富的场景表示。例如，在放射规划中，多模式数据（例如计算机断层扫描（CT）和磁共振成像（MRI）扫描）用于更精确的肿瘤轮廓，从而降低放射治疗中损伤健康组织的风险[25,29]。通常情况下，多模态传感器在模态之间自然具有不同的外部参数，例如透镜参数和相对位置。在这些情况下，非刚性图像配准对于正确执行上述下游任务至关重要。

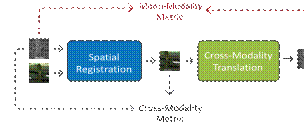


图1：方法概述。传统方法（底部虚线褪色）使用跨模态度量（例如，归一化互相关）来优化空间变换函数。我们的方法学习跨模态翻译，即两种模态之间的映射。这使得可以使用可靠、准确的单模态度量。

经典的多模态图像配准技术试图通过非线性优化过程扭曲源图像以匹配目标图像，寻求最大化预定义的相似性度量[40]。除了对自动驾驶等应用至关重要的计算缺陷外，为此类优化有效设计相似性度量已被证明是相当具有挑战性的。这适用于医学成像中常用的基于强度的测量[10]，以及基于特征的测量，通常适用于更详细的模式（例如近红外（NIR）和RGB）[32]。

这些困难导致了深度回归模型的最新发展。这些类型的模型通常有很长的训练时间，不管是有监督的还是无监督的，但是它们提供了快速的推理，通常可以很好地概括。由于为注册参数收集地面真实数据非常困难，监督多模式注册方法通常使用合成数据来训练注册网络[30,37]。这使得它们的鲁棒性高度依赖于人工和真实数据分布和外观之间的相似性。另一方面，无监督注册技术通常包含空间变换网络（STN）[14]，并训练端到端网络[7,19,16,36,8]。

通常，此类方法通过使用简单的相似性度量（例如像素均方误差（MSE））比较变形图像和目标图像来优化STN[31,33,6]。当然，这些方法只能用于单模态设置，而与多模态设置无关。为了克服这一限制，无监督多模式注册网络使用基于统计的相似性度量，特别是，（归一化）互信息（（N）MI）[22]，归一化互相关（NCC）[5]，或结构相似性指数度量（SSIM）[21,22]（见图1，虚线路径）。然而，这些指标要么难以计算（如MI）[3]，因此不能用于基于梯度的方法，要么依赖于域（如NCC），无法概括所有模式。

本文提出了一种无监督的多模态配准方法。在我们的工作中，我们利用多模态图像翻译的著名成功[13,38,39,12]，同时学习多模态翻译和空间配准。其关键思想是通过在两种给定模式上训练图像到图像的翻译网络来缓解手工制作的相似性度量的缺点。这反过来让我们可以使用单模态度量来评估注册网络（参见图1，顶部的vivid path）。*TR*

这种方法的主要挑战是同时训练注册网络和翻译网络，同时鼓励保持几何结构。这确保了这两个网络是特定于任务的-仅执行光度量映射，同时学习注册任务所需的几何变换。在我们的工作中，我们使用生成性对抗网络（GAN[9,24]）的概念来训练和控制。我们表明，对抗性训练不仅对于翻译任务是必要的（如之前的工作[13]所示），而且对于产生平滑和准确的空间变换也是必要的。我们在真实的商业数据上评估了我们的方法，并通过一系列研究证明了它的优势。我们工作的主要贡献是：*RTTTRTR*

•用于多模态图像配准的无监督方法。

•一个几何保持翻译网络，允许在多模态注册中应用单模态度量。

•鼓励生成器保持几何图形的培训计划。

# 2.  相关作品

为了处理模式之间的光测量差异，无监督的多模式方法被迫找到不同领域之间的相关性，并用它来指导他们的学习过程。在[21]中，使用香草CycleGAN体系结构来规范变形贴图。这是通过训练鉴别器网络来区分变形图像和真实图像来实现的。要对齐一对图像，整个网络需要在一次过程中进行训练。在大型数据集上训练该网络将鼓励变形映射成为身份映射。这是因为鉴别器只提供真实和变形的图像。此外，作者使用了多个跨模态相似性度量，包括SSIM、NCC和NMI，这些度量受到所用特定模态兼容性的限制。相比之下，我们的方法从一个大的数据集学习，绕过了跨模态相似性度量的需要。

Wang等人[36]试图通过学习编码器-解码器模块来创建与模态无关的特征，从而绕过领域翻译的需要。将特征反馈给STN以学习仿射变换和非刚性变换。作者使用一个简单的相似性度量（MSE）来训练他们的网络，该度量保持局部相似性，但不强制全局保真度。

在另一个极端，[8]完全依赖于对抗损失函数。他们通过将生成的注册图像提供给鉴别器网络并将其反馈用作STN的损失函数，来训练基于U网的常规STN。由于仅仅依靠鉴别器网络来指导训练，他们失去了在注册图像和目标图像之间增强局部一致性的能力。

与我们的工作最接近的是，[27]将对抗性损失与相似性度量相结合，努力在保持局部几何特性的同时正确注册图像。他们将输入编码成两个独立的嵌入，一个用于形状，一个用于内容信息，并在这些分离的嵌入上训练注册网络。这种方法依赖于学习到的解纠缠，这会在局部水平上引入不一致性。我们的方法直接在图像空间中增强相似性，从而产生可靠的局部信号。

# 3.  概述

我们的核心思想是学习两种模态之间的翻译，而不是使用跨模态度量。这种新方法如图1所示。空间变换后的图像通过可学习网络进行转换。然后，可以使用简单的单模态度量将翻译后的图像与原始源图像进行比较，而无需使用跨模态度量。使用可学习的翻译网络的优点是它可以概括和适应任何一对给定的模式。

|  |
| --- |
| 图2：培训流程概述。我们训练了两个组件：（i）空间变换网络（STN）=（RΦ，RS）和（ii）图像到图像转换网络。这两个网络通过两种不同的培训流程进行联合培训。这两个培训流程在每个培训步骤中同时执行。在第一个流中，（b）首先注册，使用由生成的变形场对输入图像进行变形，然后将其馈送至域b以将图像映射到域b。第二个流（c）首先平移，与用于变换平移的源图像的例外情况类似。*RTRTIaφR*Φ*Tφ* |

我们的注册网络由两部分组成：（i）空间转换网络=（RΦ，RS）和（ii）图像到图像的转换网络。使用两个训练流同时训练两个组件。在这两种情况下，使用相同的变形场。*RTφ*

如图2所示。空间变换网络获取两幅输入图像并生成一个变形场。然后，在该字段之前（图2b）或之后（图2c）应用该字段。具体地说，该场是使用网络生成的，并且由重采样层使用以获得变换后的图像，即（T（a），φ）和（RS（a，φ））。我们将在第4.2节详细介绍这两个培训计划。关键是，正如我们将要展示的那样，这两个流训练鼓励保持几何体，这意味着所有的几何体变换都是编码的。*φTR*Φ *RSRSTTR*Φ

一旦训练完成，测试时只使用空间变换网络。该网络以两种不同的模式拍摄两幅图像，分别代表同一场景，从稍微不同的视点拍摄，并与图像对齐。*RIaIbA.BIaIb*

# 4.  方法

我们的目标是学习一种非刚性的空间变换，它将来自不同领域的两幅图像对齐。让我们⊂ R×W×c×B⊂ R×W×Cbe两个成对的图像域，其中分别是域D的高度、宽度和通道数。配对意味着对于每个图像∈ 存在一个独特的图像∈ B表示由不同的相应传感器获取的相同场景。请注意，配对假设是一种常见且合理的假设，因为注册基应用程序通常涉及从两个模态传感器（例如卫星图像）拍摄同一场景的图像。在本节中，我们让∈ A和∈ B是一对需要对齐的两个图像*HA.A.A.HBBB高清、WD、CDIaIbIaIbIa*

*Ib*.

为了实现这种对齐，我们训练了三个可学习的组件：（i）注册网络，（ii）翻译网络和（iii）鉴别器。这三个网络使用对抗模式进行训练[9,24]，其中和被联合训练以智胜。下面，我们描述每个网络的设计和目标。*RTDRTD*

## 4.1.  注册网络

我们的注册网络（=（RΦ，RS））是一个空间转换网络（STN），由一个完全进化网络和一个再采样层组成。我们应用的变换是一种非线性密集变形，允许图像之间的非均匀映射，从而给出准确的结果。接下来，我们将对每个组件进行深入描述。*RR*Φ *RS*

**RΦ**-变形场生成器：网络获取两个输入图像，和，并生成一个变形场=R（Ia，Ib），描述如何非刚性对齐到。该场是二维向量的×WA矩阵，指示输入图像中每个像素（i，j）的变形方向。*IaIbφ IaIb哈Ia*

**RS v**-重采样层：该层接收由生成的变形场，并将其应用于源图像。这里，源图像不一定是，它可以来自域-A或B。具体地说，在像素=（i，j）处的变换图像（is，φ）的值由等式1给出：*φR*Φ*是IaRS*

*RS*（Is，φ）[v]=Is[v+φ（v）]，（1）式中（v）=(∆Y∆x） 是由at像素=（i，j）分别在和-方向上产生的变形。*φR*Φ **v***xY*

为了避免过度扭曲变形图像（Is，φ），我们限制生成非平滑变形。我们采用了一个常用的正则化项，用于生成平滑变形。特别是，正则化损失将鼓励相邻像素具有类似的变形。形式上，我们寻求一阶梯度的小值，因此像素=（i，j）处的损失由下式给出：*RSR*Φ *φ***v**

                Lsmooth（φ，v）=xb（u，v）kφ（u）− φ（v）k，（2）

**U**∈N（v）

其中（v）是像素的一组邻域，（u，v）是用于减少过平滑的双边滤波器[34]。Let=RS（Is，φ）为on输入产生的变形图像，则双边滤波器由以下公式给出：*N***v***B操作系统RS是*

*B*（紫外线，）=e−α·kO*s*[u]−O*s*[v] k*.* (3)

关于等式3中的双边滤波器，有两个重要的注释。首先，双边滤波是关于变换后的图像（在每个前向过程中），其次，项（is，u，v）被视为常数（在每个后向过程中）。后者对于避免交替像素值非常重要，以便（u，v）≈ 0（例如，它可以改变像素，使kOs[u]− Os[v]k相对较大），而前者允许更好地探索解决方案空间。*B操作系统BR*Φ *B*

在我们的实验中，我们观察了的3×3邻域，并将其设置为1。网络的整体平滑度损失，用Lsmooth（R）表示，是所有像素的平均值∈ {1，…，HA}×{1，…，WA}。**vv***α R*

## 4.2.  几何保持翻译网络

我们工作的一个关键挑战是训练图像到图像的翻译网络，使其保持几何结构。如果是几何保持，则意味着它仅执行光度量映射，因此注册任务仅由注册网络执行。然而，在我们的实验中，我们观察到，无论图像的准确性如何，都会产生与地面真实图像在空间上对齐的假图像。*TTRTR*

为了避免这种情况，我们可以通过减少容量（层数）来限制执行任何空间对齐。虽然我们确实注意到，减少的容量确实提高了注册网络的性能，但它仍然限制了注册网络完成所有注册任务（请参阅补充资料）。*TT*

为了隐含地鼓励几何保持，我们要求和是可交换的，即。，◦ R=R◦ T.在下文中，我们正式定义了两者◦ R和◦ T:*TTRTTR*

翻译优先-（R）◦ T） （Ia，Ib）：该映射首先应用图像到图像的转换，然后在转换后的图像上应用空间变换。具体地说，通过首先应用on获得最终图像，其生成假样本＝T（Ia）。然后，我们将我们的空间变换网络应用于并获得最终输出：*IaTIaOTROT*

*奥特*=RS（OT，φ）=R（T（Ia），RΦ（Ia，Ib））。

先注册—（T）◦ R） （Ia，Ib）在该构图中，我们首先对其应用空间变换，并获得变形图像=R（Ia，φ）。然后，我们使用我们的翻译网络翻译到域B：*Ia或或T*

*奥特*=T（RS（Ia，φ））=T（RS（Ia，RΦ（Ia，Ib）））。

注意，在两种成分中（即。，◦ R和◦ T） ，再采样器使用的变形场由（Ia，Ib）给出。唯一的区别是在源图像中，我们从中重新采样变形图像。*TRRSR*Φ

要理解此训练方案为什么会提供所需的属性，请注意，翻译网络提供的是几何不同的输入图像。即，这些是输入图像（在平移第一变体中）和该图像的空间变换版本（Ia，RΦ（Ia，Ib））（在配准第一方案中）。因此，鼓励保持几何图形，因为我们期望不同输入具有相似的行为。此外，请注意，它执行显式几何变换的能力有限（因为它不是指定用于此任务的）。另一方面，注册网络的设计正好相反——它自然支持几何变形，并与基于风格和外观的修改进行斗争。*TIaRSTTR*

在本节中，我们将和称为◦ T和◦ R、 分别。*奥特奥特RT*

## 4.3.  培训损失

为了训练和生成与域B中相似的假样本，我们使用1-重建损失：*RTL*

                   L（T，R）=kO− Ik1+kO− Ik1（4）*侦察RTBTRB*

如果将上述内容最小化，则意味着◦ R≈ R◦ T*T*

我们使用条件GAN（cGAN）[24]作为我们的对抗性训练损失，以及。对抗性网络的目标是区分真假样本，而和被联合训练来愚弄鉴别器。CGA的损失为◦R和◦T的公式如下：*DTRDTRTR*

LcGAN（T，R，D）=E[log（D（Ib，Ia））]

                                               + E[log（1− D（ORT，Ia））]（5）

+ E[log（1− (工程技术,建造)],，

总目标如下所示：

L（T，R）=argmaxLcGAN（T，D，R）D（6）

+λ·L（T，R）+λ·L（R），*R侦察s光滑的*

我们选择在哪里找到，这样就等于argminL（T，R）。此外，在我们的实验中，我们设置*T*∗ *R*∗ *T*∗*R*∗

*R、 T*

*λR*=100和=200。*λS*

## 4.4.  实施细节

我们的代码是使用PyTorch 1.1.0[26]实现的，并且基于Pix2Pix[13]、CycleGAN[38]和BiCycleGAN[39]的框架和实现。该网络是具有剩余连接的编码器-解码器网络[1,15]，并且注册网络是基于U-NET的[28]，在编码器和输出路径中具有剩余连接。在所有剩余连接中，我们使用实例规范化层[35]。所有网络均通过Kaiming[11]初始化方法进行初始化。*T*

实验在单GeForce RTX 2080 Ti上进行。我们使用Adam Optimizer[17]对参数为1×e的12号小批量进行优化−4、=0.5和=0.999。我们对我们的模型进行200个阶段的训练，并在100个阶段后激活线性学习率衰减。*lrβ*1 *β*2

# 5.  实验结果

在下一节中，我们将评估我们的方法，并探讨它们之间的相互作用以及我们使用的不同损失术语。*RT*

我们所有的实验都是在一个商业数据集上进行的，该数据集包含了不同生长条件和表型的香蕉植物的图像集合。该数据集包含6100个图像帧，其中每个帧由RGB图像、IR图像和深度图像组成。彩色图像是从高分辨率传感器捕获的24位彩色位图。红外图像是从长波红外（LWIR）传感器捕获的16位灰度图像。最后，由Intel Real Sense深度相机拍摄深度图像。对三个传感器进行校准，并通过深度和受控实验室测量基于仿射变换估计应用初始配准。数据集中的未对齐是由于场景中不同对象之间的深度变化造成的，这会影响初始注册

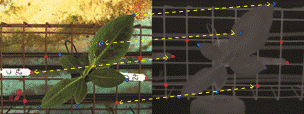


图3：注释示例。我们从源图像（左）和目标图像（右）中拾取点。蓝色点位于显著对象上，红色点是场景中的常规点。我们添加了几个箭头来说明一些匹配点。每个点的几何图形都是相对于其相应图像的。*IaIb*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| STN | 边上的SSIM | 边上的NCC | NCC | 我们的 |
| *R*仿射 | *X*/ *X* | 19.44 / 19.45 | 20.56 / 13.26 | 13.53 / 8.5 |
| *R*TPS | *X*/ *X* | 32.47 / 26.82 | 28.68 / 26.47 | 10.01 / 7.02 |
| *R* | 28.41 / 26.12 | 27.41 / 16.78 | 29.91 / 15.8 | 6.93 / 6.27 |

表1：几种相似性度量的配准精度。我们报告了不同注册网络（即，和我们的）的平均注册精度。在每个表项中，我们报告了两个精度，一个是基于完整场景注释（左）测量的，另一个是基于仅突出对象注释（右）。表示网络退化的情况。*R*仿射*R*TPS*RX*

无法处理。我们将数据集分为训练样本和测试样本，其中测试图像的采样概率为0.1。*P*

## 5.1.  评价

配准精度度量。我们手动注释了100对随机的测试图像。我们在源图像和目标图像上标记了10-15对值得注意的点地标，这些点地标预计在注册过程中匹配（见图3）。给定一对测试图像，以及一组标记对。配准网络的精度仅为目标点与其匹配变形源点之间的平均欧氏距离。*IaIbR*

此外，我们使用了两种类型的注释。第一种类型的注释位于场景中突出的对象上（图3中的蓝色点）。这一点很重要，因为在大多数情况下，下游任务主要受两种模式中场景中主要对象对齐的影响。第二个注释是通过从场景中的所有对象拾取地标点来执行的。

定量评价。由于我们工作的关键是减少对跨模态相似性度量的需求，因此我们使用常用的损失项来训练我们的网络。在表1中，我们展示了我们的注册网络在使用不同方法进行培训时的注册精度-

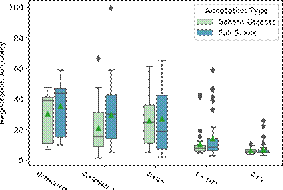


图4：不同方法的配准精度。未注册表示数据集中的未对齐。我们展示了通过Elastix[18,23]和基于SIFT[20]和CycleGAN[38]或我们的翻译网络（即）的基于特征的注册获得的准确性。*T*

耳鼻喉科指标。具体来说，我们使用了归一化互相关（NCC）度量，因为它经常用于无监督的多模态配准方法。此外，我们通过最大化Canny边缘检测器[4]从变形图像和目标图像检测到的边缘上的相似性度量（例如，结构相似性指数度量（SSIM）和NCC）来训练我们的网络。从表1可以看出，使用规定的跨模态相似性度量训练注册网络的效果不佳。此外，使用NCC会产生有噪声的结果，而使用SSIM会产生平滑但不太准确的配准精度（参见补充资料）。*R*

此外，由于我们学习的度量是通用的，并且可以与任何空间变换合并，因此我们也使用我们的度量训练不同的空间变换。具体来说，我们训练了基于仿射的STN（）和基于薄板样条的STN（）。表1中报告了这些网络的注册精度。从表中可以看出，使用我们的度量对这些网络进行训练，与其他损失项相比，会产生显著的改进。*R*仿射*R*TPS

我们还将我们的方法与三种技术进行了比较。我们考虑的第一种方法是SimpleLastix[18,23]，这是一种基于互信息的迭代注册技术。另外两种方法是基于特征的技术，其中SIFT[20]描述符适用。然而，由于SIFT[20]不是为多模态数据设计的，因此它不能直接用于源图像和目标图像。

相反，我们训练CycleGAN[38]网络在手边的两种模式之间进行转换，而不需要任何监督来匹配基本事实。CycleGAN与其他无监督图像到图像的翻译网络一样，没有经过训练以生成与地面真值样本匹配的图像，因此，翻译网络不明确要求进行几何变换。训练后，我们使用CycleGAN中的一个生成器（将域A映射到域B的生成器）将输入图像转换到模态B上。假设该生成器既保持了几何结构，又在模态之间进行了良好的转换，预计它也能很好地匹配假样本和目标图像的特征。因此，我们通过CycleGAN翻译网络从生成的图像中提取SIFT描述符，并从目标图像中提取SIFT特征。然后我们匹配这些特征并估计所需的空间配准。与直接在输入图像上使用SIFT[20]特征相比，使用该方法的配准精度显著提高。同样，我们使用了我们的几何保持翻译网络以及SIFT描述符。*IaIbIaT*

在图4中，我们展示了上述方法的准确性。可以看出，我们的方法具有最佳的配准精度。此外，与CycleGAN[38]+SIFT[20]和Elastix[23,18]相比，使用我们的翻译网络和SIFT描述符实现了实质性的改进。因此，我们的转换器既能保持几何结构，又能进行精确的光测量映射。*TT*

定性评价。图5显示，我们的注册网络成功地将来自不同模式对的图像对齐，并处理不同的对齐情况。例如，图5a中第一个未加工的香蕉叶在两种模式中很好地对齐。我们的注册网络保持这种对齐，并且只变形背景以实现图像之间的完全对齐。这可以从变形场可视化[2]中看出，其中对香蕉植物应用的变形较小，而大部分变形应用于背景。此外，在图5a的第二行中，由于摄像机的移动，大部分图像在某个方向上平移，但是仍然可以看到与深度相关的变化。为了帮助测量对齐成功率，我们在配准前后将图像B中的植物覆盖在图像A的顶部（使用半透明度）。这意味着轮廓在所有图像（原始图像B、配准前后的图像A）中具有相同的空间位置。最后，我们在RGB和IR图像之间的配准方面取得了类似的成功（见图5b）。

值得一提的是，在某些情况下，变形场指向源图像之外的区域。在这些情况下，我们只需对零值进行采样。这是因为这些区域中的目标图像内容（即）在源图像（即）中不可用（有关更多结果，请参阅补充资料）。*IbIa*

## 5.2.  烧蚀研究

|  |
| --- |
| 输入A输入B注册变形。前后场  （a） RGB和深度模式之间的图像配准。                  输入A输入B注册变形。前后场  （b） RGB和IR模式之间的图像配准。  图5：定性评价。我们展示了两对域之间注册的示例结果；（a） RGB到深度注册和（b）RGB到红外注册。在前两列中，我们显示了相应的图像和。第三列是注册图像，即变形后的图像。变形场（第4列）使用标准光流可视化[2]进行可视化。最后，我们分割突出的对象，并在配准前后（最后两列）在相同的空间位置将其覆盖（不透明度为25%）。*IaIbIaIb* |

接下来，我们将介绍一系列消融研究，分析我们工作中不同方面的有效性。首先，我们证明了训练两个组成部分（即我们提出的两个训练流）确实能够鼓励一个几何保留译者。此外，我们还分析了不同损失项对注册网络准确性的影响。我们进一步证明了双边滤波的有效性，它确实提高了配准精度。除非另有说明，否则所有实验均在不进行双边过滤的情况下进行。几何保留翻译网络。为了评估图2中提出的两个训练流的训练和同时训练的影响，我们比较了我们的方法与训练模型的配准精度◦ R或◦ T.如图6所示，与每个训练流（即。，◦ R和◦ T） 分开。而且，在重建的同时，损失了◦ R（以read显示）是三个选项中最低的，它不一定表示更好的注册。这是因为在此设置中，翻译网络隐式执行对齐和翻译任务。相反地，当使用◦仅T（显示为绿色），网络不稳定，在某个点开始交替像素值，基本上承担翻译网络的角色。由于只有几何图形由设计感知，因此无法生成良好的样本。这通过鉴别器检测到生成的样本是假的速度来表示（即，对抗性损失衰减得很快）。补充材料中提供了视觉效果。*TRTTRTRTRTTRRR*

损失消融。在以前的工作[39,13,38]中已经表明，训练图像到图像的翻译网络，同时进行重建和对抗损失，可以产生更好的结果。特别是，重建损失稳定了训练过程并提高了输出图像的逼真度，而对抗性损失则鼓励生成与真实数据分布相匹配的样本。

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 图6：成分消融研究。我们展示了（a）配准精度，（b）重建损失，（c）对抗性损失和（d）cGAN损失的值。所有图形中的x轴都是历元数。显示的损耗值适用于以下情况：◦ R（红色），◦ T（绿色）和我们的（蓝色）。可以看出，使用我们的方法配准精度最好。在里面◦ R、 重建损失最低，但是，注册不准确，因为注册任务的很大一部分是由翻译人员隐式执行的。此外，组成◦T是不稳定的，因为在某一点上，注册网络开始交替像素值，这是由鉴别器检测到的（参见（c）中的倾斜）。*TRTTRR*   |  |  |  |  |  | | --- | --- | --- | --- | --- | |  | |  |  |  | | --- | --- | --- | | 方法 | 测试依据。 | 列车附件。 | | | |  |  |  | | --- | --- | --- | | 没有注册 | 35.45 | 34.96 | | 不带双边 | 11.01 | 9.89 | | 双边 | 6.93 | 6.12 | | |

我们工作的主要目标是建立一个登记网络。因此，我们试图了解表2：烧蚀损失结果。列表示使用GAN损耗项训练的模块。行表示使用L1损失项训练的模块。我们不报告只有一个模块未接受任何损失条款培训的结果。表示训练偏离的情况。例如，第二行和第一列中的结果表示当模块的权重相对于cGAN损失和模块相对于重建损失项更新时达到的配准精度。*XRT*

损失（重建和对抗）对注册网络的影响。为了了解每个损失的影响，我们使用不同的设置来训练我们的模型：每次我们固定或的权重，相对于其中一个损失函数。配准精度如表2所示。定性结果请参考补充材料。从这些图中可以看出，仅针对重建损失进行训练会导致过于尖锐但不现实的图像，其中变形场会产生噪声伪影。另一方面，仅针对敌方损失进行的训练会产生逼真的图像，但不精确对齐。这在表2中尤其明显，其中关于重建损失的训练实现了对齐的显著改善，并且当损失项都用于更新所有网络权重时，获得了最佳精度。*RTRRR*

双边过滤有效性使用双边过滤来衡量平滑度损失，实际上，允许我们鼓励变形贴图上的分段平滑度。如表3所示，这提高了配准的精度。这些结果表明，使用分段表3：平滑正则化。双边滤波对配准精度的影响。我们展示了带注释的测试样本和带注释的序列样本的配准精度。

用于控制平滑度损失项的偏移图可能是有益的。

# 6.  摘要和结论

提出了一种基于图像到图像转换网络的无监督多模态图像配准技术。具体来说，我们开发了一个几何保持的图像到图像转换网络，该网络允许使用简单的单峰度量来比较变形图像和目标图像。通过一种新的训练方案，交替并组合两种不同的流来训练空间变换，使得几何保持翻译网络成为可能。

我们相信，几何保持生成器可以用于图像配准以外的应用。在未来，我们将继续探索在不同的流中交替训练几个层或操作符的想法，以鼓励它们作为实现某些非平凡属性的手段是可交换的。

# 致谢

这项研究得到了以色列创新局和Phenomics财团的通用研发项目的支持。

# 工具书类

[1] 深度残差学习用于图像识别。2016年IEEE计算机视觉和模式识别会议（CVPR），第770-778页，2015年。5.

[2] 西蒙·贝克、丹尼尔·沙尔斯坦、J.P.刘易斯、斯特凡·罗斯、迈克尔·布莱克和理查德·塞利斯基。光流数据库和评估方法。国际计算机杂志。《愿景》，92（1）：2011年3月1日至31日。6, 7

[3] Mohamed Ishmael Belghazi、Aristide Baratin、Sai Rajeshwar、Sherjil Ozair、Yoshua Bengio、Aaron Courville和Devon Hjelm。互信息神经估计。詹妮弗·迪和安德烈亚斯·克劳斯主编，《第35届国际机器学习会议录》，机器学习研究会议录第80卷，第531-540页，瑞典斯德哥尔摩Stockholmsmssan，2018年7月10-15日。PMLR。2.

[4] J坎尼。边缘检测的计算方法。IEEE Trans。肛门型。机器。《国际法》，第8（6）：679-6981986年6月。6.

[5] 曹晓欢、杨建环、李旺、钟雪、钱旺、沈定刚。基于深度学习的模式内相似性监督的多式联运图像配准。医学影像学中的机器学习。MLMI，11046:55–632018。2.

[6] 阿德里安诉达尔卡、古哈·巴拉克里希南、约翰诉古塔格和默特·R·萨布奇。用于快速概率微分同胚配准的无监督学习。2018年，米奇。2.

[7] 鲍勃·德沃斯、弗洛里斯·F·贝伦森、马克斯·A·维杰弗、马吕斯·斯泰格和伊凡娜·伊斯古姆。采用卷积神经网络进行端到端无监督可变形图像配准。在M.Jorge Cardoso、Tal Arbel、Gustavo Carneiro、Tanveer F.Syeda Mahmood、Joao Manuel R.S.Tavares、Mehdi Moradi、Andrew P.Bradley、Hayit Greenspan、Joao Paulo Papa、Anant Madabushi、Jacinto C.Nascimento、Jaime S.Cardoso、Vasileios Belagiannis和Zhi Lu中，编辑，医学图像分析深度学习和临床决策支持多模式学习-第三届国际研讨会，DLMIA 2017，第七届国际研讨会，ML-CDS 2017，与MICCAI 2017联合举办，加拿大魁北克市，QC，2017年9月14日，会议记录，第10553卷计算机科学讲稿，第204-212页。斯普林格，2017年。1.

[8] 范景凡、曹晓欢、王倩、叶佩钦、沈定刚。针对单模式或多模式注册的对抗式学习。医学图像分析，58:10154522019。1, 2

[9] 伊恩·古德费罗、让·普吉·阿巴迪、迈赫迪·米尔扎、徐冰、大卫·沃德·法利、谢尔吉尔·奥扎尔、艾伦·库尔维尔和约书亚·本吉奥。生成性对抗网。在Z.Ghahramani、M.Welling、C.Cortes、N.D.Lawrence和K.Q.Weinberger的《神经信息处理系统的进展》27中，编辑，第2672-2680页。科伦联营公司，2014年。2, 3

[10] Grant Haskins、Uwe Kruger和Pingkun Yan。医学图像配准的深度学习：调查，2019年。1.

[11] 何开明、张向宇、任少清、孙健。深入研究整流器：在imagenet分类方面超越人类水平的性能。2015年IEEE国际计算机视觉会议（ICCV）会议记录，ICCV&apos;15，第1026-1034页，美国华盛顿特区，2015年。IEEE计算机协会。5.

[12] 黄迅、刘明宇、贝隆吉和考茨。多模态无监督图像到图像的转换。在ECCV，2018年。2.

[13] 菲利普·伊索拉、朱俊彦、周廷辉和阿列克谢·埃弗罗斯。使用条件对抗网络进行图像到图像的翻译。CVPR，2017年。2, 5, 7

[14] Max Jaderberg、Karen Simonyan、Andrew Zisserman和Koray Kavukcuoglu。空间变压器网络。在Corinna Cortes、Neil D.Lawrence、Daniel D.Lee、Masashi Sugiyama和Roman Garnett主编的《神经信息处理系统的进展28：2015年神经信息处理系统年会》，2015年12月7日至12日，加拿大魁北克省蒙特利尔，2017-2025页，2015年。1.

[15] 贾斯汀·约翰逊、亚历山大·阿拉希和李菲菲。实时风格转换和超分辨率的感知损失。2016年欧洲计算机视觉会议。5.

[16] 金宝儿、金杰恩、李琼、金东焕、朴成浩和钟楚烨。使用循环一致CNN的无监督可变形图像配准。沈定刚、刘天明、Terry M.Peters、Lawrence H.Staib、Caroline Essert、Sean Zhou、Pew Thian Yap和Ali Khan，医学图像计算和计算机辅助干预-MICCAI 2019-第22届国际会议编辑，中国深圳，2019年10月13日至17日，《计算机科学讲稿》第六部分第11769卷，第166-174页。斯普林格，2019年。1.

[17] 迪德里克·金马和吉米·巴。Adam：随机优化的一种方法。学习表现国际会议，2014年12月12日。5.

[18] 斯蒂芬·克莱因、马吕斯·斯泰德、基林·墨菲、马克斯·A·维尔杰和乔森·P·W·普鲁姆。elastix：用于基于强度的医学图像配准的工具箱。IEEE Trans。医学。影像学，29（1）：196-2052010。6.

[19] 马修·C·H·李、奥桑·奥克泰、安德烈亚斯·舒赫、米切尔·夏普和本·格洛克。用于结构引导图像配准的图像和空间变换网络。沈定刚、刘天明、Terry M.Peters、Lawrence H.Staib、Caroline Essert、Sean Zhou、Pew Thian Yap和Ali Khan，医学图像计算和计算机辅助干预-MICCAI 2019-第22届国际会议编辑，中国深圳，2019年10月13日至17日，会议记录，第二部分，第11765卷计算机科学讲稿，第337-345页。斯普林格，2019年。1.

[20] 大卫·G·洛。从比例变化的关键点中提取独特的图像特征。国际计算机杂志。《愿景》，60（2）：91-1102004年11月。6.

[21]德瓦里卡纳特·马哈帕特拉、巴夫纳·安东尼、苏曼·塞代和拉希尔·加尔纳维。使用生成对抗网络的可变形医学图像配准。第15届IEEE生物医学成像国际研讨会，2018年ISBI，美国华盛顿特区，2018年4月4-7日，第1449-1453页。IEEE，

2018. 2

[22]德瓦里卡纳特·马哈帕特拉、宗元·葛、苏门·塞代和拉吉布·查克拉沃蒂。使用生成对抗网络对X射线图像进行联合配准和分割。《医学影像学中的机器学习》编者Shi Yinghuan、Xuang Il-Suk和Liu Mingxia，计算机科学课堂讲稿第11046卷，第73-80页。斯普林格，2018年1月1日。2.

[23]卡斯珀·马斯塔尔、弗洛里斯·F·贝伦森、马吕斯·斯泰格和斯特凡·克莱因。SimpleLastix：一个用户友好的多语言医学图像注册库。2016年6月26日至7月1日，在美国内华达州拉斯维加斯举行的2016年IEEE计算机视觉和模式识别研讨会、2016年CVPR研讨会上，第574-582页。IEEE计算机学会，2016年。6.

[24]Mehdi Mirza和Simon Osindero。条件生成对抗网。CoRR，abs/1411.17842014。2, 3, 4

[25]吴承正和金思勇。放射治疗中的变形图像配准。放射肿瘤学杂志，35（2）：1012017。1.

[26]Adam Paszke、Sam Gross、Soumith Chintala、Gregory Chanan、Edward Yang、Zachary DeVito、Zeming Lin、Alban Desmaison、Luca Antiga和Adam Leer。Pytork中的自动微分。在NIPS Autodiff研讨会上，2017年。5.

[27]陈琴、史碧波、廖瑞、托马索·曼西、丹尼尔·鲁克特和阿里·卡门。通过解纠缠表示实现多模态图像的无监督可变形配准。艾伯特C.S.C.胡，Paul A. Yushkevich，司琦葆，编辑，医学影像信息处理-第二十六国际会议，IPMI 2019，香港，中国，六月2-7，2019，程序，第11492卷，计算机科学讲义，第249页至261页。斯普林格，2019年。2.

[28]O.Ronneberger、P.Fischer和T.Brox。U-net：用于生物医学图像分割的卷积网络。《医学图像计算和计算机辅助干预》（MICCAI），LNCS第9351卷，第234-241页。斯普林格，2015年。（可在arXiv上获得：1505.04597[cs.CV]）。5.

[29]玛丽亚A施密特和杰弗里S佩恩。利用mri进行放射治疗计划。《医学与生物学物理学》，60（22）：R323，2015年。1.

[30]N.Schneider、F.Piewak、C.Stiller和U.Franke。Regnet：使用深度神经网络的多模态传感器注册。2017年IEEE智能车辆研讨会（IV），第1803-1810页，2017年6月。1.

[31]A.Sheikhjafari、Michelle Noga、Kumaradevan Punithakumar和Nilanjan Ray。基于全连通生成神经网络的无监督可变形图像配准。20182.

[32]沈晓勇、李旭、张琪、贾佳娅。自然图像的多模态和多光谱配准。在计算机视觉-ECCV 2014-第13届欧洲会议中，

*瑞士苏黎世，2014年9月6日至12日，诉讼记录，第四部分*，第309-324页，2014年。1.

〔33〕常树、陈曦、齐伟勰、华翰。用于快速显微图像配准的无监督网络。John E.Tomaszewski和Metin N.Gurcan主编，《2018年医学影像：数字病理学》，第10581卷，第363页

– 370. 国际光学与光子学学会，SPIE，2018年。2.

[34]C.Tomasi和R.Manduchi。灰度和彩色图像的双边滤波。《第六届计算机视觉国际会议记录》，ICCV&apos;98，第839页–美国华盛顿特区，1998年。IEEE计算机协会。4.

[35]Dmitry Ulyanov、Andrea Vedaldi和Victor S.Lempitsky。实例规范化：快速样式化缺少的要素。ArXiv，abs/1607.08022016。5.

[36]王成家、乔戈斯·帕帕纳斯塔西奥、阿吉西劳斯·沙特西斯、格泽戈兹·贾森科、索蒂里奥斯·沙夫塔里斯、张和野。FIRE：使用深度网络的无监督双向模态间配准。CoRR，abs/1907.0506219。1, 2

[37]阿尔芒·赞皮耶里、纪尧姆·夏皮亚特、尼古拉斯·吉拉德和尤利娅·塔拉巴尔卡。通过多尺度神经网络链进行多模态图像对齐，并应用于遥感。在ECCV，2018年。1.

[38]朱俊彦、朴大成、菲利普·伊索拉和阿列克谢·埃夫罗斯。使用cycleconsistent敌对网络进行未配对图像到图像的转换。计算机视觉（ICCV），2017年IEEE国际会议，2017年。2, 5, 6, 7

[39]朱俊彦、张泽楷、迪帕克·帕萨克、特雷弗·达雷尔、阿列克谢·埃夫罗斯、奥利弗·王和伊莱·谢赫特曼。多模态图像到图像的转换。神经信息处理系统的进展，2017年。2, 5, 7

[40]芭芭拉·齐托夫和扬·弗卢瑟。图像配准方法综述。图像与视觉计算，21（11）：977-1000，

2003. 1