

## 【论文笔记】Spatial Transformer Networks

原创

2016年07月02日 12:46:04

14785

参考文献: \*\*Jaderberg M, Simonyan K, Zisserman A. *Spatial transformer networks*[C]//*Advances in Neural Information Processing Systems*. 2015: 2017-2025.

### 摘要

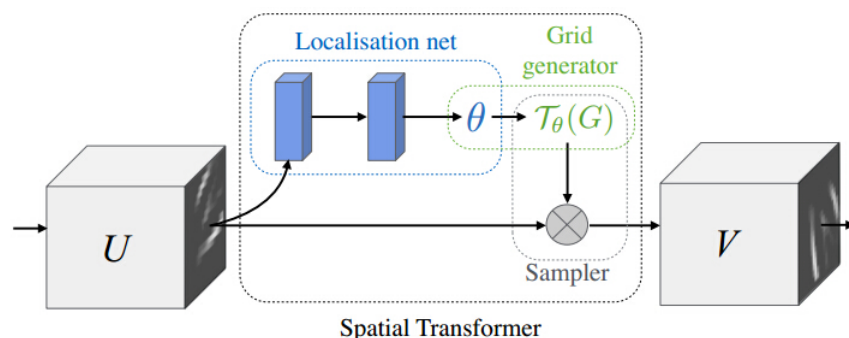
卷积神经网络 (CNN) 已经被证明能够训练一个能力强大的分类模型, 但与传统的模式识别方法类似, 它也会受到数据在空间上多样性的影响。这篇Paper提出了一种叫做空间变换网络 (**Spatial Transform Networks, STN**), 该网络不需要关键点的标定, 能够根据分类或者其它任务自适应地将数据进行空间变换和对齐 (包括平移、缩放、旋转以及其它几何变换等)。在输入数据在空间差异较大的情况下, 这个网络可以加在现有的卷积网络中, 提高分类的准确性。

由于我之前的工作部分涉及到人脸对齐, 所以看到这篇Paper异常激动。总觉得能用它做点什么。

### 算法介绍

#### 1. 算法总流程

STN 主要可以分为三个部分: 1) localisation network. 2) grid generator. 3) sampler. (中文我翻译不准确, 大家意会下)。localisation network用来计算空间变换的参数  $\theta$ , grid generator则是得到input map  $U \in R^{H \times W \times C}$  到 output map 各位置的  $V \in R^{H' \times W' \times C}$  对应关系  $\mathcal{T}_\theta$ , sampler根据input map  $U$  和 对应关系  $\mathcal{T}_\theta$ , 生成最终的output map. 流程图如图所示:



#### 1.1 Localisation Network

它的作用就是通过一个子网络 (全连接或者卷积网, 再加一个回归层), 生成空间变换的参数  $\theta$ 。  $\theta$  的形式可以多样, 如需实现2D仿射变换,  $\theta$  就是一个6维 (2x3) 向量的输出。

#### 1.2 Parameterised Sampling Grid

假设  $U$  (不局限于输入图片, 也可以是其它层输出的feature map) 每个像素的坐标为  $(x_i^s, y_i^s)$ ,  $V$  的每个像素坐标为  $(x_i^t, y_i^t)$ , 空间变换函数  $\mathcal{T}_\theta$  为仿射变换函数, 那么  $(x_i^s, y_i^s)$  和  $(x_i^t, y_i^t)$  的对应关系可以写为:

$$\begin{pmatrix} x_i^s \\ y_i^s \end{pmatrix} = \mathcal{T}_\theta(G_i) = A_\theta \begin{pmatrix} x_i^t \\ y_i^t \end{pmatrix}$$

当然,  $A_\theta$  也可以有其它形式, 如3D仿射变换, 透射变换等。

#### 1.3 Differentiable Image Sampling

在计算得到  $\mathcal{T}_\theta$  后, 就可以由以下公式  $U$  得到  $V$  了 (省略推导公式若干, 只放上最终形式):

$$V_i^c = \sum_n^H \sum_m^W U_{nm}^c \max(0, 1 - |x_i^s - m|) \max(0, 1 - |y_i^s - n|)$$

在求得  $V$  后, 当然少不了上述公式对  $U, x^s, y^s$  的求导, 以便根据loss进行网络的后向传播:

$$\frac{\partial V_i^c}{\partial U_{nm}^c} = \sum_n^H \sum_m^W \max(0, 1 - |x_i^s - m|) \max(0, 1 - |y_i^s - n|)$$
$$\frac{\partial V_i^c}{\partial x_i^s} = \sum_n^H \sum_m^W U_{nm}^c \max(0, 1 - |x_i^s - m|) \max(0, 1 - |y_i^s - n|) \begin{cases} 0, & \text{if } |m - x_i^s| \geq 1 \\ 1, & \text{if } m \geq x_i^s \\ -1 & \text{if } m < x_i^s \end{cases}$$

$\frac{\partial V_i^c}{\partial y_i^s}$  与  $\frac{\partial V_i^c}{\partial x_i^s}$  类似。对  $\theta$  的求导为:

$$\frac{\partial V_i^c}{\partial \theta} = \begin{pmatrix} \frac{\partial V_i^c}{\partial x_i^s} \cdot \frac{\partial x_i^s}{\partial \theta} \\ \frac{\partial V_i^c}{\partial y_i^s} \cdot \frac{\partial y_i^s}{\partial \theta} \end{pmatrix}$$

而  $\frac{\partial x_i^s}{\partial \theta}, \frac{\partial y_i^s}{\partial \theta}$  根据具体的变换函数便可得到。

通过以上3个部分的结合, 便形成了完整的  $STN$ 。

2. 算法分析

$STN$  计算较快, 几乎没有增加原有网络模型的训练时间。由于它能够在训练过程中, 学习到与任务相关的空间变换参数, 因此能够进一步最小化网络的损失函数。 $STN$  不只可以用在输入的图像层, 也可以加入卷积层或者其它层之后。

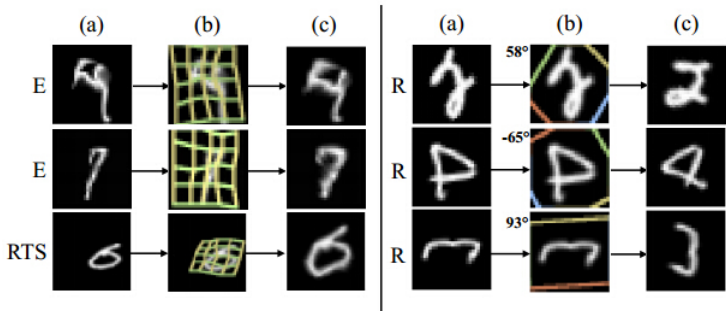
3. 实验结果

这篇文章分别在手写文字识别、街景数字识别、鸟类分类以及共定位等方面做了实验, 这里我只列出比较有代表性的手写文字实验部分。

实验数据为MNIST, 分别在经过不同处理 (包括 旋转 ( $R$ )、旋转、缩放、平移 ( $RTS$ ), 透射变换( $P$ ), 弹性变形 ( $E$ )) 的数据上进行字符识别的实验。Baseline分别使用了两种网络结构 $FCN, CNN$ , 加入了  $STN$  的网络为  $ST-FCN, ST-CNN$ 。其中,  $STN$  采用了以下几种变换方法: 仿射变换 ( $Aff$ )、透射变换 ( $Proj$ )、以及薄板样条变换 ( $TPS$ )。下表列出了  $STN$  与  $baseline$  在MNIST上的比较结果, 表中数据为识别错误率:

Model		MNIST Distortion			
		R	RTS	P	E
FCN		2.1	5.2	3.1	3.2
CNN		1.2	0.8	1.5	1.4
ST-FCN	Aff	1.2	0.8	1.5	2.7
	Proj	1.3	0.9	1.4	2.6
	TPS	1.1	0.8	1.4	2.4
ST-CNN	Aff	0.7	0.5	0.8	1.2
	Proj	0.8	0.6	0.8	1.3
	TPS	0.7	0.5	0.8	1.1

可以看出, 对不同的形式的数据, 加入了 $STN$  的网络均优于  $baseline$  的结果。以下为  $STN$  对数字图像进行变换后的结果, 其中a列为原始数据, b列为变换参数的示意图, c列为最终变换后的结果:



## 总结

STN 能够在没有标注关键点的情况下，根据任务自己学习图片或特征的空间变换参数，将输入图片或者学习的特征在空间上进行对齐，从而减少物体由于空间中的旋转、平移、尺度、扭曲等几何变换对分类、定位等任务的影响。加入到已有的CNN或者FCN网络，能够提升网络的学习能力。

版权声明：本文为shaoxiaohu原创文章，欢迎转载，请注明出处，谢谢。

<https://blog.csdn.net/shaoxiaohu1/article/details/51809605>

本文已收录于以下专栏：[CV论文笔记](#)



### 从小白到AI工程师的学习经验分享

这是转型AI的励志故事，从非科班到拿下阿里云栖一等奖，他经历的坑足够你学习100天！以下为他的正文分享，你可以清晰地看到他趟过的每一个坑，希望借他的肩，让你勇敢前行。

12522

[查看更多>>](#)

👤 目前您尚未登录，请 [登录](#) 或 [注册](#) 后参与评论

## 深度学习方法（十二）：卷积神经网络结构变化——Spatial Transformer Networks

今天具体介绍一个Google DeepMind在15年提出的Spatial Transformer Networks，相当于在传统的一层Convolution中间，装了一个“插件”，可以使得传统的卷积...



xbinworld 2017年04月03日 23:45 🗨️ 7316

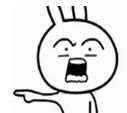
### 空间映射网络--Spatial Transformer Networks zhangjunhit 2017年03月28日 09:29 🗨️ 2032

Spatial Transformer Networks 主要对目标在特征空间做不变性归一化 解决 角度、尺度等变形引入的影响 Code: <https://github.com/skaae/t...>

### 码农怎能不懂英语？！试试这个数学公式

老司机教你一个数学公式秒懂天下英语

广告



### 【论文笔记】Spatial Transformer Networks shaoxiaohu1 2016年07月02日 12:46 🗨️ 14785

卷积神经网络（CNN）已经被证明能够训练一个能力强大的分类模型，但与传统的模式识别方法类似，它也会受到数据在空间上多样性的影响。这篇Paper提出了一种叫做空间变换网络（Spatial Transfo...

### 论文阅读《Spatial Transformer Network》 yaoqi\_isee 2017年05月27日 22:01 🗨️ 618

Google DeepMindAbstract:作者说明了CNN对于输入的数据缺乏空间变换不变形(lack of spatially invariant ability to input data), ...

### 论文笔记：Spatial Transformer Networks（空间变换网络）

上一篇博客 Spatial Transformer Networks论文笔记（一）——仿射变换和双线性插值介绍了仿射变换和双线性插值，为更好地理解STN打基础。本篇博客是记录的是阅读原文Spatial...



sinat\_34474705 2017年07月17日 21:18 🗨️ 1203

### 知网论文查重入口

知网论文检测入口在哪

百度广告



## Tensorflow1.0空间变换网络(SpatialTransformer Networks)实现

空间变换网络简单介绍：通过locatnet，提取输入图像的theta(将用于仿射变换)；根据输入图像的width和height以及仿射变换(或者TPS)的参数theta，可以生成目标位...



weixin\_36368407 2017年03月01日 10:01 1459

## Paper Reading: Spatial Transformer Networks (with code explanation)

原文: Spatial Transformer Networks前言: 卷积神经网络(CNN)已经可以构建出一个强大的分类模型，但它仍然缺乏能力来应对输入数据的空间变换，。这篇文章提出了一种叫做空间变换网...



l691899397 2016年12月14日 17:25 2118

## [深度学习论文笔记][Attention] Spatial Transformer Networks

Jaderberg, Max, Karen Simonyan, and Andrew Zisserman. "Spatial transformer networks." Advances in Ne...



Hao\_Zhang\_Vision 2016年11月15日 22:02 1303

## 论文阅读《Spatial Transformer Networks》



zzchust 2015年11月09日 12:26 5169

STN module inserted to CNN without any extra training feature maps learn invariance to translation, ...

## 空间映射网络--Spatial Transformer Networks



u014568921 2017年05月31日 21:54 690

深度学习方法(十二): 卷积神经网络结构变化——Spatial Transformer Networks 空间映射网络--Spatial Transformer Networks...



对于程序员来说，英语到底多重要  
不背单词和语法，一个公式学好英语

## A brief scanning of paper "Spatial Transformer Network"

Part.1: what's the problems the method proposed in this paper solved? When we using a convolutio...



sinat\_24002967 2017年11月09日 22:06 60

## Spatial Transformer Network学习笔记



u014229172 2017年12月11日 21:16 127

学习了MatConvNet的STN的代码，网络结构是这样的：affinegrid和sampler都是没有要更新的weight的，有ST的训练结果是这样：没有ST的训练结果是这样：...

## STN系列之Spatial Transformer Networks



dreamer\_on\_air 2017年09月04日 13:32 457

Spatial Transformer Networks. Max Jaderberg Karen Simonyan Andrew Zisserman Koray KavukcuogluGoogle ...

## [深度学习论文笔记][CVPR 17 oral] Inverse Compositional Spatial Transformer...

这篇文章是针对Spatial Transformer Network进一步改进的工作。从研究领域来看，该工作是对增强深度网络之于输入图片空间不变性的研究。...



u010158659 2017年04月23日 00:44 6542

## 2016-Spatial Transformer Networks理解

2018年01月17日 10:00 680KB

下载



## 知网论文查重入口

在知网上查重论文一般需要多长时间

百度广告



## 详细解读Spatial Transformer Networks (STN) -一篇文章让你完全理解STN了

目录 STN的作用 1.1 灵感来源 1.2 什么是STN? STN的基本架构 Localisation net是如何实现参数的选取的? 3.1 实现平移 3.2 实现缩放 3.3 实现旋转...



qq\_39422642 2017年12月22日 10:36 693

---

## Spatial Transformer Networks



keyanxiaocaicai 2017年06月11日 21:06 245

caffe-stn移植到高版本的方法 参考: <http://blog.csdn.net/kuaitoukid/article/details/51035028> <https://github.com/ha...>

---

## Spatial Transformer Networks(空间变换神经网络)

闲扯: 大数据不如小数据 这是一份很新的Paper(2015.6), 来自于Google旗下的新锐AI公司DeepMind的四位剑桥Phd研究员。他们针对CNN的特点, 构建了一个新的局部...



brandon2015 2017年05月12日 17:56 1260

---

## 空间变换网络--spatial transform network



u011961856 2017年09月10日 11:11 2189

CNN分类时, 通常需要考虑输入样本的局部性、平移不变性、缩小不变性, 旋转不变性等, 以提高分类的准确度。这些不变性的本质就是图像处理的经典方法, 即图像的裁剪、平移、缩放、旋转, 而这些方法实际上就是对图像...

---

## Spatial Transformation Network



DuinoDu 2017年04月07日 12:46 877

STN

---