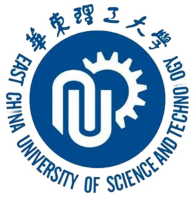



中国研究生创新实践系列大赛

“华为杯”第十七届中国研究生数学建模竞赛

学　　校　郑州大学

参赛队号　20104590039

队员姓名　1.张璇
　　2.刘晓强
　　3.彭文斌

目录

中国研究生创新实践系列大赛

“华为杯”第十七届中国研究生数学建模竞赛

题 目 能见度估计与预测

摘 要：

当今社会衣食住行，出行对人们的生活可谓是非常的重要，而影响交通出行的一个重要因素便是能见度。能见度差往往意味着恶略天气，而恶略天气又极易导致交通事故的发生，从而严重影响交通运输条件，给我们的出行造成困扰，极易形成严重的交通事故，对人们的生命和财产造成损失。近些年来，雾霾天气在我国各大城市愈演愈劣，如何对交通运输条件进行准确的判断就非常的重要了，其中最为重要的便是准确快速的进行能见度检测。

传统的大气能见度检测方法分为人眼目测法和仪器测量法，人眼目测法主要是指：在人眼没有任何帮助的条件下，所能识别物体的最大距离[1][2]，这往往会引入很多的人为误差，给测量结果带来不确定性，而且人眼目测客观性相对较差；仪器测量法主要是指：使用一些光学器件来对大气条件中的能见度值进行检测，这种设备通常比较昂贵，且在雨、雾等低能见度天气中，会因水汽吸收等复杂条件造成较大的测量误差。

针对本项目问题一，本文利用某机场气象测量点监测到的能见度、气压、温度等相关数据，用 Pearson 相关系数法找出影响能见度的主要因子，并构建能见度与气象学因子的模型，用多元回归分析对模型求解，得到两类不同的能见度模型。经过比较，两类模型均能很好的符合能见度观测数据，可以利用模型做短时预测。最后，通过能见度与各因子的散布分析，验证了模型的准确性，并给出能见度与湿度、温度、风速、气压之间的经验关系：在低温、低压、底风速、高湿度的情况下，低能见度的概率较高。

针对本项目问题二，本文针对机场视频数据设计了一种基于改进的卷积神经网络（CNN）的雾霾能见度检测方法，首先根据视频流数据截取对应图片，然后得到机场交通雾霾图像库，然后分为训练数据集和测试数据集，将训练数据集首先进行分类，按照能见度的不同动态的划分为 9 类，之后进行预处理然后送入深度卷积神经网络提取雾霾特征向量，最后根据总的特征向量利用 Softmax 函数进行分类输出，得到每一类能见度值的概率，最终输出概念值最大的能见度值。本文在分析现有数据的情况下，对分类做了动态的划分，将能见度较为低的情况划分了足够的类别，最终在有限条件下取得了 77.8%的模型识别率。

针对问题三、四，图像中的能见度估计，本文使用大气散射模型将能见度的求取问题转化为大气消光系数的求取问题，而大气消光系数可以利用暗通道先验理论求出大气透射率和几何条件进行建模求出目标距离代入公式求出。利用这些信息创建了基于特定目标点透射率的能见度估计模型，通过对图片进行暗通道处理和去雾处理得到透射率图和去雾后的图像，在去雾后的较清晰图像中准确定位设定目标点，找到该点的透射率值，并根据交比不变性原理找到灭点和灭线，求出目标点到观测点即摄像机的距离，最后代入经推导得到的能见度公式求出能见度估计值。通过对问题三、问题四求解出的能见度估计和预测进行分析，发现估计的能见度值在实际允许的可靠范围内，预测的能见度变化趋势与实际变化趋势相符。

关键字：气象学因子，多元回归分析，机器视觉，卷积神经网络，暗通道先验

1 问题重述

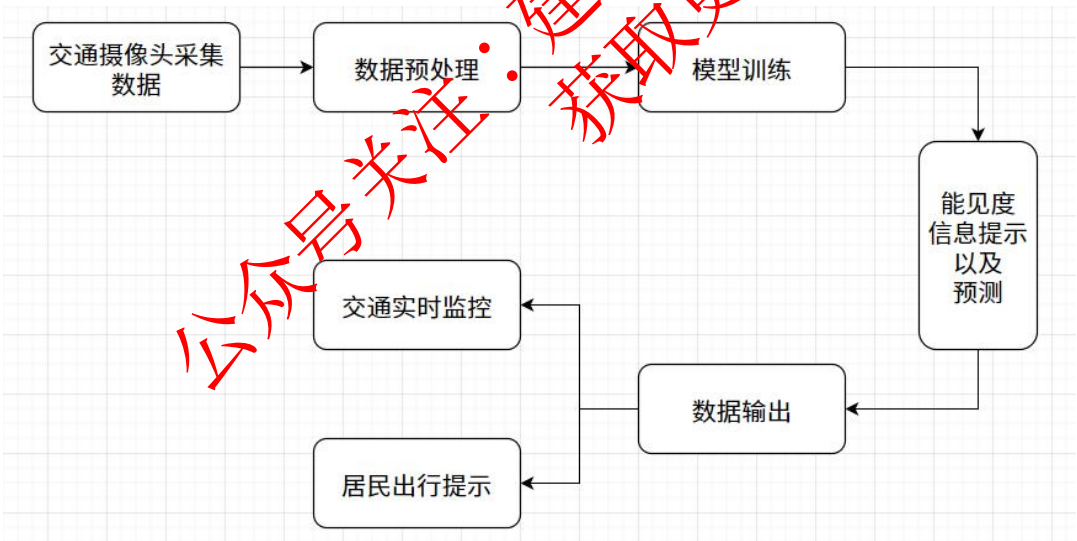
1.1 问题背景

伴着世界经济的高速发展，近些年来很多重大交通事故的发生原因都是由于雾天能见度过低而导致的。大气能见度也代表着我们所生活的环境中空气质量的好坏，能见度较低代表空气中的污染物颗粒以及悬浮颗粒物的含量较高，对人体伤害较大，雾霾天气能见度低在对我国经济发展产生危害的同时，也严重的影响到了我国人民的生命和财产安全。

虽然我们无法避免恶劣天气因素的发生，但对危险情况的提前预测同样可以带来很大的安全保障，尤其是对雾天能见度的监测，可以极大程度的减少交通事故发生的概率。因此对于雾天能见度检测与预测方法的研究，不仅有利于人们的身体健康，还能够为人们的生活出行安全提供重要导向[3]。随着我国汽车保有量的不断飙升，当我们出门开车行驶的时候最需要关注的就是对大气能见度的及时检测和准确预测。

以往能见度检测主要有人工法和仪器法，人工法缺乏重复性和客观性，仪器法因为仪器采购昂贵且需要后期培训来使用并且往往伴随着后期对仪器的维护工作，因此难以大范围推广。随着科学技术的不断发展，利用机器视觉技术配合神经网络技术便可以解决上述的两个难点，在我国交通路网中，遍布的交通监控摄像头可以提供大量的有效的雾霾数据，交通摄像头采集数据之后可以配合后续计算机技术来完成对数据的处理。针对本问题设计的能见度估计系统流程图如下图 1 所示。

图 1 能见度估计系统流程图



1.2 问题的分析

在本项目中，有以下 3 个问题需要解决。

根据所提供的机场气象观测数据，寻找能见度与气象因素（温度、湿度、风速等）之间的关系，建立预测模型。

根据某机场的视频数据和对应的能见度数据，先找出视频数据与能见度数据之间的关系，并依据推断得出的关系设计一个深度学习模型，使得视频数据和能见度数据之间的能够得到匹配，并利用数据评估模型的准确程度。

高速公路的监控视频只有视频数据，而无与能见度观测仪一样的数字观测数据，要想获知视频数据中的能见度信息，需根据不同能见度下物体的亮度和景深，设计算法处理自视频中截取的图像，计算得到每张图片场景的能见度，分析能见度随时间的变化规律。利用问题三得到的能见度随时间变化规律，建立数学模型预测大雾变化趋势，如何时加重或减弱、何时散去等。

1.3 问题的分析

依据以上提出的问题，我们作出了如下分析。

本题中已知能见度、风速、气压、湿度等数据，根据气象数据得到等能见度预测模型，属于一个因变量、多个自变量的多元模型。首先进行数据预处理，对冗杂数据进行筛选，得到有效数据；然后在有效数据中，对剩余自变量进行Pearson相关分析，减少相似数据的干扰，得到5-7个有效变量；再对有效变量分别与因变量做相关分析，判断每个自变量与因变量的相关程度，筛选3-5个对因变量影响较大的因素；之后对剩余变量建立回归模型，得到预测模型；最终检验模型准确性。

在本题中得到了机场视频数据和某段连续时间能见度数据表。首先要做的是找到视频数据和能见度数据之间的对应关系，两者的数据量在数量上应该一致。之后便可以设计深度学习网络，深度学习网络设计的最终目的是输出一个能见度数据或者输出能见度在某个范围内的概率最大。在训练网络时，将上述匹配好的机场数据与能见度数据分为两部分，一部分作为训练数据集另一部分作为验证数据集。

本题目给出的信息只有 100 张截取自高速公路监控视频的图片，图片中显示该张图片对应的实际时间，图中还隐藏有道路标线的尺寸等信息。通过 Koschmieder 的能见度检测理论，对图像进行数字图像处理，计算得到能见度检测算法需要的各项数据，代入能见度计算公式，即可求出图片中场景的能见度。对得到的不同时间点能见度信息进行分析，建立能见度随时间变化模型，预测未来大雾的变化趋势。

2 模型假设与符号说明

2.1 模型假设

为了使本文所设计模型在尽量接近实际环境的前提下，能够合理简化分析过程，在建模时提出如下假设。

所有数据均为原始数据，来源真实可靠。

假设在机场视频数据中，能见度变化情况在一分钟内不会发生剧烈变化。也即在机场视频数据中某一分钟内的某张或者某几张图片的数据可以代表当前分钟内的能见度的数据。

假定在能见度数据中，每分钟的数据是间隔 15 秒进行采集。

假定视频数据在分类完成之后，如果出现肉眼可见的与实际采集数据不匹配的问题，选择忽略此图片。

视频截图中高速公路的车道分界线（虚线）符合国家标准，长 6 米，间隔 9 米。

高速公路视频监控摄像头的高度为 6 米，相对位置在拍摄期间保持不动。

2.2 名词解释

RVR：跑道视程，是指在跑道中线上的航空器上的飞行员能看到跑道面上的标志或跑道边界灯或中线灯的距离。

MOR：光学视程，是指色温为 2700K 的白炽灯发出的平行光束的光通量在大气中衰减降低到它的起始值的 5%的距离。

能见度（白天）：指的是在白天条件下，观测者将有足够亮度的目标物从周围环境中识别出来的最远距离。在气象学范畴中，气象能见度定义为在不借助任何辅助设备的条件下，观测者可以从水平视线范围内识别出大小合适的黑色目标物的最大距离。

2.3 符号说明

本文中所使用的符号有如下的意义。

表格 1 符号说明

符号	意义
Conv	卷积层
Pool	池化层
Fc	全连接层
CNN	卷积神经网络
Softmax	Softmax 逻辑回归模型
ReLu	线性整流
R^2	可决系数
Sig.	差异性显著检验值
MRA	多元回归分析

3 问题的模型建立与求解

3.1 问题一模型的建立与求解

3.1.1 问题一的描述

已有研究发现，一些由于天气因素造成的航空事故中，有近 20%是由于低能见度造成的。因此，根据已知观测数据，建立能见度与气象因子的关系，掌握其变化规律，能有效地对能见度进行预测，保障飞行安全。在本题中，已知能见度等各类气象因子，求解模型属于多元回归模型。分析各变量间相关性，找到能见度的有效影响因子，能大大优化模型。

3.1.2 问题一的分析与建立

根据世界气象组织的定义，能见度是指正常视力的人在白天当时的天气条件下，能够从天气背景中看到和辨认出适当大小黑色目标物的最大距离，在夜间则指假设亮度与白天相同条件下能够辨认出目标物的最大距离，其好坏程度会受到温度、湿度、风和降水等气象因素以及地形作用、人类活动等非气象因素的影响，有关研究表明[4]，大气能见度的优劣与气象条件密切相关，尤其是相对湿度和地面风速。比如日出之前，温度较低，空气中湿度较大，此时对于光线传播产生不利影响，因此能见度会降低，晨雾就是这类情况；日出之后，温度逐渐回升，湿度变小，空气中水分较少，气压升高，能见度也变大；此外，风速也是影响能见度的重要因素：风速变大，空气中气溶性物质、雾霾等移动速度变快，此时能见度将会增加。

MRA是数据处理中重要一个方法，针对有多个变量时，可以将相关变量中一个变量视为因变量，其余多个变量视为自变量，建立多个变量之间线性或非线性的数学模型与数量关系式，对样本数据进行分析。对于多元回归分析，重要的一步就是对自变量进行相关分析，得到与因变量相关性较高的的因子，这样在建模过程中才能建立比较合适的模型。因此，选择温度、湿度、风速、风向、垂直风速、露点温度、站点压力、最高点压力等主要气象因素作为自变量，建立与能见度MOR之间的关系。

(1)相关性分析：分析变量间的相关性一个重要方法就是 Pearson 相关系数法，相关性系数是介于[-1，+1]之间的实数。当相关性系数小于 0 时，表明变量之间存在负相关关系；当相关性系数大于 0 时，表明变量之间存在正相关关系；当相关性系数为 0 时，二者之间不存在相关性。同时，相关性系数绝对值越接近 1，表明变量之间的相关性越强，当相关系数越接近 0，表明变量之间的相关性越弱。

由于研究对象的不同，相关系数有多种定义方式，其中最基本的定义式为：

$$r(X,Y)=\frac{Cov(X,Y)}{\sqrt{Var[X]Var[Y]}} \tag{1}$$

其中，$Cov(X,Y)$为 X 与 Y 的协方差，$Var[X]$、$Var[Y]$分别为 X、Y 的协方差。当相关性系数的绝对值介于 0.1~0.3 之间时，一般认为变量间存在弱相关；当相关系数的绝对值介于 0.3~0.5 之间时，一般认为变量间存在中度相关；当相关性系数的绝对值大于 0.5 时，一般认为变量间存在强相关。本文中，为拟合得到最佳模型，认为变量间相关系数大于 0.7 为强相关，存在相关性；若自变量间相关性达到 0.9 以上，则认为两变量可以互相代替，选择其一即可。

Pearson相关系数能准确的反应两变量之间是否相关，但对变量之间有一定限制：数据为连续变量且成对出现；数据无异常值；两组数据之间呈线性相关关系；数据服从正态分布。

当变量间满足以上条件时，即可用Pearson相关系数法求解变量相关性，若不满足，可

对数据预处理（如剔除异常值，非线性转化为线性等）[5]。

(2)多元回归分析：回归分析是一种处理变量的统计相关关系的一种数理统计方法，多元回归分析是研究多个变量之间关系的回归分析方法。其基本思想是：虽然自变量和因变量之间没有严格、确定的函数关系，但是可以设法找出最能代表他们之间关系的数学表达式。它能解决以下几类问题：确定几个特定的变量之间是否存在相关关系，如果存在的话，找出它们之间合适的数学表达式；根据一个或几个变量的值，预测或控制另一个变量的取值；进行因素分析。在对于共同影响一个变量的许多变量之间，找出因素之间重要程度。

判断回归拟合程度的指标称为拟合优度（Goodness of Fit），度量拟合优度的统计量是可决系数（亦称确定系数）R^2。R^2最大值为 1。R^2的值越接近 1，说明回归直线对观测值的拟合程度越好；反之，R^2的值越小，说明回归直线对观测值的拟合程度越差[6]。其计算方式为：

设 y 为待拟合数值，其均值为$\bar{y}$,拟合值为$\hat{y}$,记总平方和（SST）：$\sum_{i=1}^{n}(y_i-\bar{y})^2$，回归平方和（SSR）：$\sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2$，残差平方和（SSE）：$\sum_{i=1}^{n}(y_i-\bar{y}_i)^2$，则有 SST=SSR+SSE，可决系数。

$$R^2=\frac{SSR}{SST}=\frac{\sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2}{\sum_{i=1}^{n}(y_i-\bar{y})^2}=1-\frac{SSE}{SST} \quad (2)$$

3.1.3 问题一的求解

模型一的求解总共分为以下五个部分，下面分别对每一个部分进行叙述。

第一个部分为：数据说明及预处理。材料中所提供的机场MOS观测数据，包含不同时间段内两天的观测数据，分别为2019年12月16日、2020年3月13日两天的观测数据，每份数据由气象观测仪器给出相关气象学指标。

具体数据内容包括：20191216：站点数据 1：记录时间、站点编号、气压、温度、相对湿度等。站点数据 2：能见度 RVR、能见度 MOR 等。站点数据 3：风速、风向、垂直风向等。其中站点数据 1 的采样频率为 1 次/分钟，采样周期为 2019 年 12 月 15 日 08:00-2019 年 12 月 16 日 07:59 共 24 小时，站点数据 2、站点数据 3 的采样频率为 4 次/分钟，采样周期为 2019 年 12 月 15 日 08:00-2019 年 12 月 16 日 07:59 共 24 小时。特别地，在站点数据 2、3 中，由于数据存储较大，出现数据丢失的情况：9：26、16：00、18：35、20：21、7：18，这五分钟均有一次数据丢失，共丢失 5 次。

20200313：记录内容与采样频率和上述相同，采样周期为 2020 年 03 月 12 日 08:00-2020 年 03 月 13 日 07:59 共 24 小时，特别地，同样由于数据存储较大，在站点数据 2、3 中，也有数据丢失的情况：9：12、18：05、21：29、1：39、4：11，这五分钟均有一次数据丢失，共丢失 5 次。

为计算数据模型，需将站点数据1与站点数据2、3进行统一，我们选择将2、3的数据压缩：每分钟四次采样取均值作为该分钟的输出。另外，对丢失数据进行人工筛查，为避免数据突变过大，丢失数据选择同一分钟第三次的采样，补充为第四次采样。

人工补充数据后，使用MATLAB将站点数据2、3压缩至每分钟采样一次，并挑选指定数据（时间、能见度RVR、能见度MOR、风速、风向、垂直风速、站点气压、海平面气压、最高点气压、温度、相对湿度、露点温度）建立新数据集。

第二个部分为：自变量Pearson相关分析，选择20191216的预处理后的数据集进行分析，使用SPSS数据分析软件对数据进行相关分析，本文假设自变量之间线性相关关系（一些非线性可以转化为线性关系的如1/X，认为是线性的），暂不考虑复合类型关系。为得到各变量之间的相关系数，需要判断数据是否能利用Pearson分析，即看各变量之间是否满足上述条件，以温度与露点温度两个变量为例。

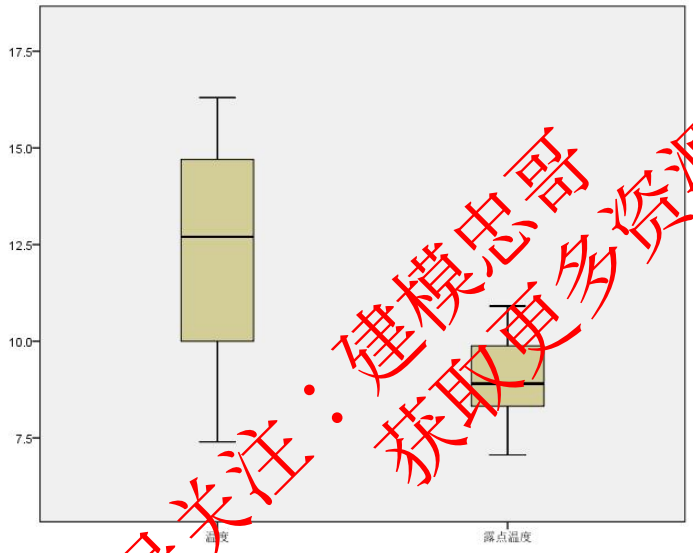
数据为连续变量且成对出现；温度与露点温度数量成对且为连续变量。

数据无异常值；数据异常值对相关性分析的影响较大，检验数据是否存在异常值我们一般在SPSS中绘制箱形图来完成。步骤入下：将两个变量绘制箱型图，如图2所示，可以看到，本次变量1440个，全部纳入图形绘制，无缺失数据；没有数据点位于箱图误差线的上下范围，因此这两个变量的数据没有异常点。

图 2 温度与露点温度箱型图

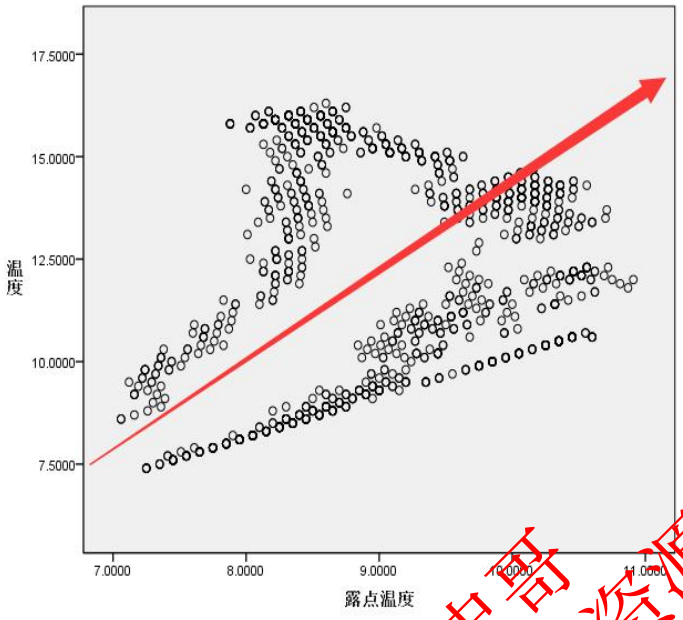
案例处理摘要

	案例					
	有效		缺失		合计	
	N	百分比	N	百分比	N	百分比
温度	1440	100.0%	0	.0%	1440	100.0%
露点温度	1440	100.0%	0	.0%	1440	100.0%



两组数据之间呈线性相关关系；本文利用SPSS绘制散点图来判断两数据见是否满足线性关系。如图3所示，可以看到温度与露点温度近似分布在红线两侧，可以认为是满足线性相关，但相关程度较弱。

图 3 温度与露点温度散点分布图



数据服从正态分布；变量的Pearson相关性分析的前提条件是数据正态分布，在SPSS中也提供了相应的功能：可以看到显著性水平较差，均小于0.05，不满足正态分布，两者是非线性相关。或者说，线性相关性较差。

表格 2 正态性检验

项目	统计量	df	Sig.	统计量	df	Sig.
露点温度	0.118	1440	0.000	0.953	1440	0.000
温度	0.132	1440	0.000	0.922	1440	0.000

其余变量的判断方式与上述步骤相同，通过对变量之间的先验检测，可大概知道自变量间的线性相关与，此时再将所有自变量进行 Pearson 相关性分析，步骤如下：导入数据；分析；相关；双变量；

各自变量间的相关系数矩阵如下图所示，由于系数矩阵是对称阵，因此只需关注上半矩阵。比较变量间相关系数可知以下关系：站点气压、最高点气压，海平面气压三个变量间高度相关，相关系数最高为1.000，三变量间彼此可以代替；露点温度与其他因素相关性较差，与气压呈负相关关系；温度与相对湿度呈高度负相关，这与经验相同：温度升高，空气水分蒸发较快，相对湿度降低；

通过分析可以得出，站点气压、最高点气压，海平面气压三气压值高度相关，可用站点气压代替其余两个气压变量，减少变量冗余。

第三个部分为：因变量与自变量相关分析。在上一步中，通过相关分析将高度相关的自变量同化，使自变量个数减少到7个，有效降低数据冗余度。在这一步我们将考虑因变量，将因变量作为一个输入变量，获得因变量与其他变量的相关性，为提高模型的准确率，在这一步还将考虑非线性的相关，即引入Spearman相关系数。对相关系数比较发现，输入的7个变量中，露点温度与能见度的关系很弱，Pearson相关性数、Spearman相关系数均不超过0.5；风向因素同样对能见度的影响较小，Pearson相关性数仅有0.3。因此可以认为这两个变量对能见度的影响有限，建模回归时可以不考虑。

图 4 自变量间相关系数矩阵

相关系数

			风速	风向	垂直风速	站点气压	最高点气压	海平面气压	温度	相对湿度	露点温度
Spearman 的 rho	风速	相关系数	1.000	.599**	.883**	.752**	.752**	.758**	.715**	-.764**	-.048
		Sig.（双侧）	.	.000	.000	.000	.000	.000	.000	.000	.068
		N	1440	1440	1440	1440	1440	1440	1440	1440	1440
	风向	相关系数	.599**	1.000	.854**	.578**	.578**	.583**	.580**	-.574**	.057*
		Sig.（双侧）	.000	.	.000	.000	.000	.000	.000	.000	.031
		N	1440	1440	1440	1440	1440	1440	1440	1440	1440
	垂直风速	相关系数	.883**	.854**	1.000	.753**	.753**	.759**	.725**	-.741**	.017
		Sig.（双侧）	.000	.000	.	.000	.000	.000	.000	.000	.528
		N	1440	1440	1440	1440	1440	1440	1440	1440	1440
	站点气压	相关系数	.752**	.578**	.753**	1.000	1.000**	.999**	.587**	-.723**	-.188**
		Sig.（双侧）	.000	.000	.000	.	.000	.000	.000	.000	.000
		N	1440	1440	1440	1440	1440	1440	1440	1440	1440
	最高点气压	相关系数	.752**	.578**	.753**	1.000**	1.000	.999**	.587**	-.723**	-.188**
		Sig.（双侧）	.000	.000	.000	.	.	.000	.000	.000	.000
		N	1440	1440	1440	1440	1440	1440	1440	1440	1440
	海平面气压	相关系数	.758**	.583**	.759**	.999**	.999**	1.000	.595**	-.731**	-.196**
		Sig.（双侧）	.000	.000	.000	.000	.000	.	.000	.000	.000
		N	1440	1440	1440	1440	1440	1440	1440	1440	1440
	温度	相关系数	.715**	.580**	.725**	.587**	.587**	.595**	1.000	-.946**	.160**
		Sig.（双侧）	.000	.000	.000	.000	.000	.000	.	.000	.000
		N	1440	1440	1440	1440	1440	1440	1440	1440	1440
	相对湿度	相关系数	-.764**	-.574**	-.741**	-.723**	-.723**	-.731**	-.946**	1.000	.101**
		Sig.（双侧）	.000	.000	.000	.000	.000	.000	.000	.	.000
		N	1440	1440	1440	1440	1440	1440	1440	1440	1440
	露点温度	相关系数	-.048	.057*	.017	-.188**	-.188**	-.196**	.160**	.101**	1.000
		Sig.（双侧）	.068	.031	.528	.000	.000	.000	.000	.000	.
		N	1440	1440	1440	1440	1440	1440	1440	1440	1440

**. 在置信度（双侧）为 0.01 时，相关性是显著的。
*. 在置信度（双侧）为 0.05 时，相关性是显著的。

表格 3 相关系数

		能见度 MOR	风速	风向	垂直风速	站点气压	温度	相对湿度	露点温度
能见度 MOR	Pearson 相关性数	1	.770	.313	.735	.814	.811	-.869	-.067
	Sig.（双侧）		.000	.000	.000	.000	.000	.000	.011
	N	1440	1440	1440	1440	1440	1440	1440	1440
能见度 MOR	Spearman 相关系数	1.000	.817	.586	.780	.873	.766	-.885	-.257
	Sig.（双侧）		.000	.000	.000	.000	.000	.000	.000
	N	1440	1440	1440	1440	1440	1440	1440	1440

第四个部分为：在上一步中，通过将因变量与自变量联合分析，使自变量个数减少到5个，因此，在这一步中可以对因变量进行曲线拟合（在SPSS中进行），得到因变量与自变量间的变化趋势。这里我们不妨提出以下假设：因变量（能见度MOR）与自变量之间呈线性关系，即可以表示为。

$$Y = aX_1 + bX_2 + cX_3 + dX_4 + \cdots + nX_n + C \quad (3)$$

其中X_1、X_2…X_n为不同自变量，a、b、c…n为对应的系数，C为常数项。同时我们假设一些非线性如X^2、1/X等都可通过简单变换变为线性（不包括复合类型）。有以上假设，可以认为在其他因素保持不变的情况下，单因子变化趋势能反应因变量的变化。本文可以进行因变量与各个自变量之间分别曲线拟合具体步骤如下：分析；回归；曲线估计；选择因变量（MOR）；选择自变量；选择回归模型。

依次选则自变量分别与因变量进行曲线回归，各个变量的回归模型，统计该自变量在不同模型下的R^2。R^2能反应回归直线对观测值的拟合程度。R^2最大值为1，R^2的值越接近1，说明回归曲线对观测值的拟合程度越好；反之，R^2的值越小，说明拟合程度越差。如表所

示，最终得到能见度与剩余五个变量间各种拟合模型的拟合程度。结果见下表。

表格 4 能见度——风速各回归模型汇总和参数估计值

	模型汇总					参数估计值			
	R 方	F	df1	df2	Sig.	常数	b1	b2	b3
线性	.593	2097.123	1	1438	.000	-732.133	1488.075		
对数	.562	1843.529	1	1438	.000	-76.671	3947.283		
倒数	.384	894.531	1	1438	.000	6653.894	-5970.711		
二次	.614	1142.309	2	1437	.000	-2344.261	2593.568	-160.085	
三次	.661	931.579	3	1436	.000	1718.392	-2293.808	1417.357	-147.834
复合	.372	851.305	1	1438	.000	171.902	2.115		
幂	.374	859.263	1	1438	.000	223.075	2.048		
S	.281	561.335	1	1438	.000	8.956	-3.248		
增长	.372	851.305	1	1438	.000	5.147	.749		
指数	.372	851.305	1	1438	.000	171.902	.749		
Logistic	.372	851.305	1	1438	.000	.006	.473		

表格 5 各因素不同模型 R^2

	线性	对数	倒数	二次	三次	复合	幂	S	增长	指数	Logistic
风速	0.5932	0.5618	0.3835	0.6139	0.6606	0.3719	0.3740	0.2808	0.3719	0.3719	0.3719
垂直风速	0.5407	.*	.**	0.6089	0.6166	0.3214	.*	.**	0.3214	0.3214	0.3214
站点气压	0.6618	0.6627	0.6635	0.6618	0.6618	0.5088	0.5098	0.5108	0.5088	0.5088	0.5088
温度	0.6583	0.6734	0.6770	0.6769	0.6769	0.4681	0.4872	0.4978	0.4681	0.4681	0.4681
相对湿度	0.7550	0.7147	0.6704	0.8701	0.8808	0.5252	0.4835	0.4405	0.5252	0.5252	0.5252

(注：*. 自变量(垂直风速)包含非正数值。最小值为-0.9725。无法计算对数模型和幂模型。

**. 自变量(垂直风速)包含零值。无法计算倒数模型和S模型。)

根据以上内容，能将模型简化到5个自变量，同时得到各自变量与因变量的大致关系。此时即可通过最小二乘法原理用MATLAB中regress函数实现。依据表2得到的模型R^2，将自变量风速、垂直风速、站点气压、温度、相对湿度定义为X_1-X_5，回归采用两种模型，分别是线性模型Y1与非线性模型Y2：

$$Y1 = aX_1 + bX_2 + cX_3 + dX_4 + eX_5 + C \quad (4)$$

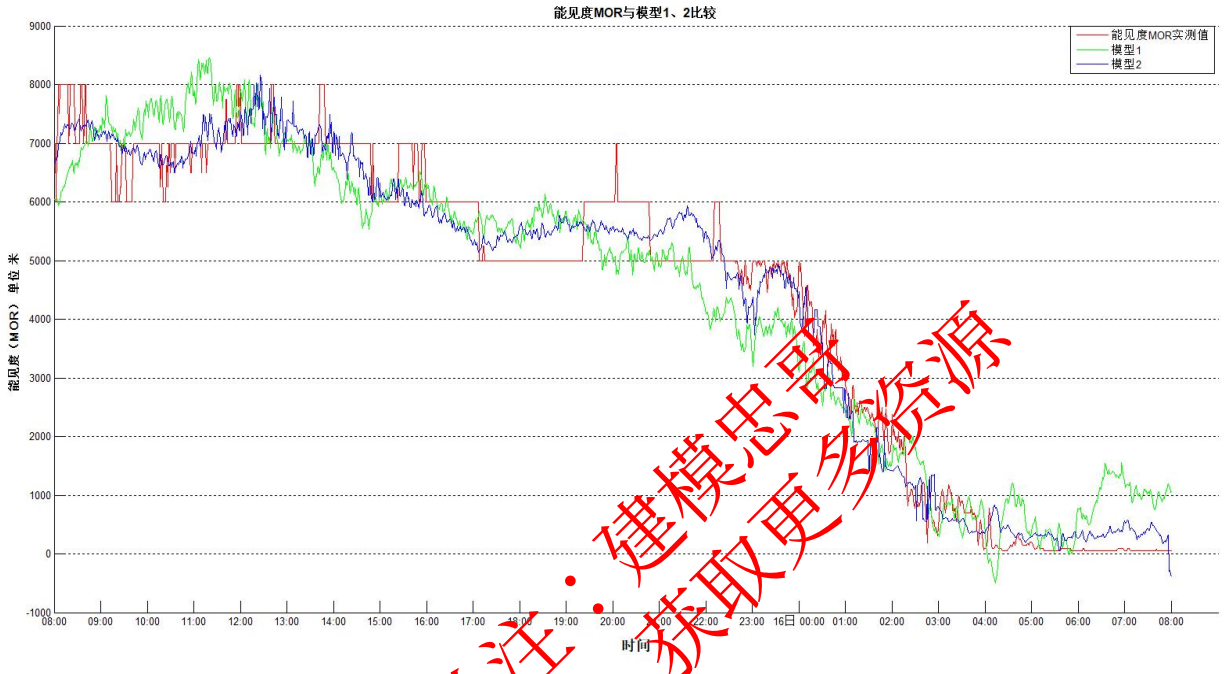
$$Y2 = (a_1X_1{}^3 + a_2X_1{}^2 + a_3X_1) + (b_1X_2{}^3 + b_2X_2{}^2 + b_3X_2) + c_1\frac{1}{X_3} + d_1\frac{1}{X_4} + (e_1X_5{}^3 + e_2X_5{}^2 + e_3X_5) + C \quad (5)$$

通过模型回归，两种模型的R^2分别为0.9168、0.9679，Sig.均小于0.01，说明两者的拟合程度都很好。最终的回归方程为公式（5）、公式（6），将原始能见度、模型1、模型2绘制在一起如下图所示。

$$Y1 = -48.168 * X_1 - 304.171 * X_2 + 940.724 * X_3 + 595.670 * X_4 - 7.247 * X_5 - 960320.482 \quad (6)$$

$$Y2 = -18.732 * X_1{}^3 + 216.95 * X_1{}^2 - 523.33 * X_1 + 6.48 * X_2{}^3 - 83.35 * X_2{}^2 + 137.92 * X_2 - 405994468 * \frac{1}{X_3} + 1069.23 * \frac{1}{X_4} - 0.45 * X_5{}^3 + 104.21 * X_5{}^2 - 8001.32 * X_5 + 607613.60 \quad (7)$$

图 5 能见度 MOR 与模型 1、2 比较



对拟合得到的公式与曲线图分析可得到以下结论：拟合得到的公式虽然很复杂，但模型1、模型2能大致反应原始能见度曲线；拟合曲线中，两个模型均为经验公式，只能做到短时预测，如两个模型虽与实际能见度走势相同，并不代表实际能见度与气象因子的关系；从模型的图和R^2可以看出，模型2相比模型1有更好的拟合结果，特别是在末尾处，模型一已经偏离较多。

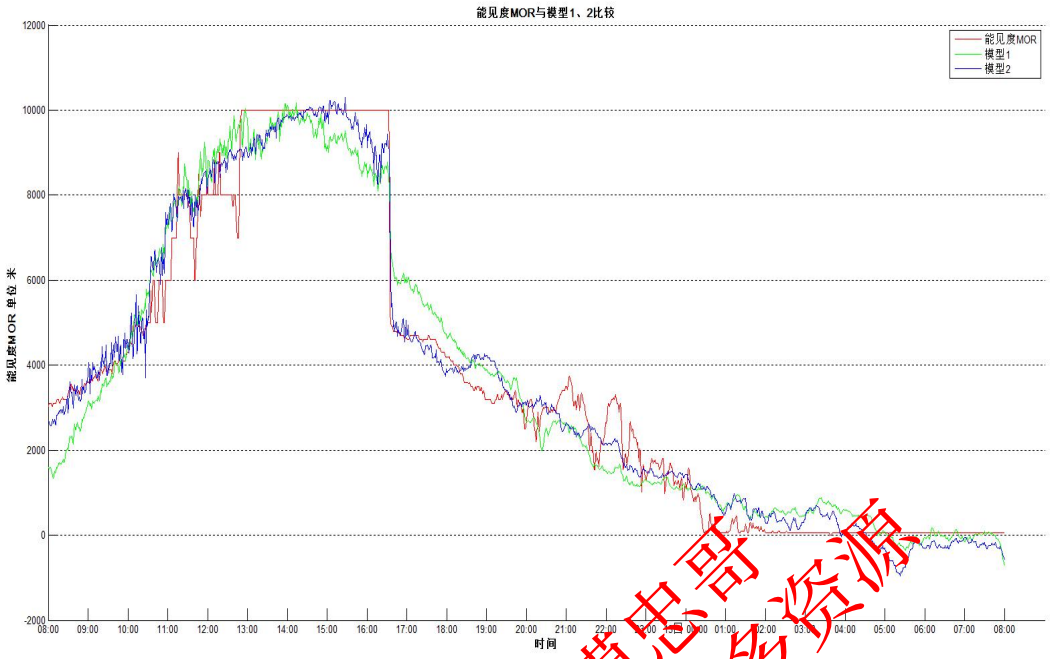
以上以20191216数据为例，从原始数据中通过相关分析，回归分析得到较为准确的能见度与风速、湿度、气压等气象学因子的关系，为能见度的短时预测提供可能。

第五个部分为：20200313数据处理。由于两天时间跨度较大，数据采集时的季节不同，两组数据的气象因素有较大不同。对第二批数据用同样的方法分析，进行数据预处理、相关分析、模型预测、模型回归，得到最终的拟合公式及模型图。本次数据根据模型将风速、露点温度、气压、温度、相对湿度定义为X_1-X_5。

$$Y1 = -249.26 * X_1 - 589.06 * X_2 - 518.56 * X_3 - 837.37 * X_4 + 65.1 * X_5 + 511963 \quad (8)$$

$$Y2 = -6.28 * X_1{}^3 - 14.61 * X_1{}^2 + 125.03 * X_1 + 10.48 * X_2{}^3 - 284.81 * X_2{}^2 + 17627.25 * X_2 + 678383117.41 * \frac{1}{X_3} + 6.81 * X_4{}^3 - 288.41 * X_4{}^2 - 10997.2 * X_4 - 0.26 * X_5{}^3 + 83.7 * X_5{}^2 - 11129.66 * X_5 - 150871.2 \quad (9)$$

图 6 能见度 MOR 与模型 1、2 比较（0313 数据）



第六个步骤为：气象因子散布分析。我们将能见度≤1Km称为低能见度，统计低能见度时相对湿度、温度等气象因子的分布，有利于我们验证回归模型并给出能见度与湿度、温度的变化趋势。

以1216数据为例，绘制能见度与相对湿度、风速、温度的散点图。通过图7、8与表4分析可知，低能见度时，相对湿度往往很高，从图7可以看出，能见度3Km以下时，相对湿度高达95%以上，且95%能见度均值仅有787m。分析风速与能见度的关系亦可以看出，低能见主要集中在4m/s以下，风速大于4m/s时的能见度都在4000米以上；在温度-能见度和气压-能见度散点图中有着类似的结论：低能见度主要集中在温度、气压相对较低的区域。

表格 6 各档相对湿度对应能见度均值

	相对湿度			
	≥95%	90%-94%	85%-89%	<80%
能见度均值（m）	787	4622	5704	6280

图 7 相对湿度-能见度关系（左）和风速-能见度关系（右）

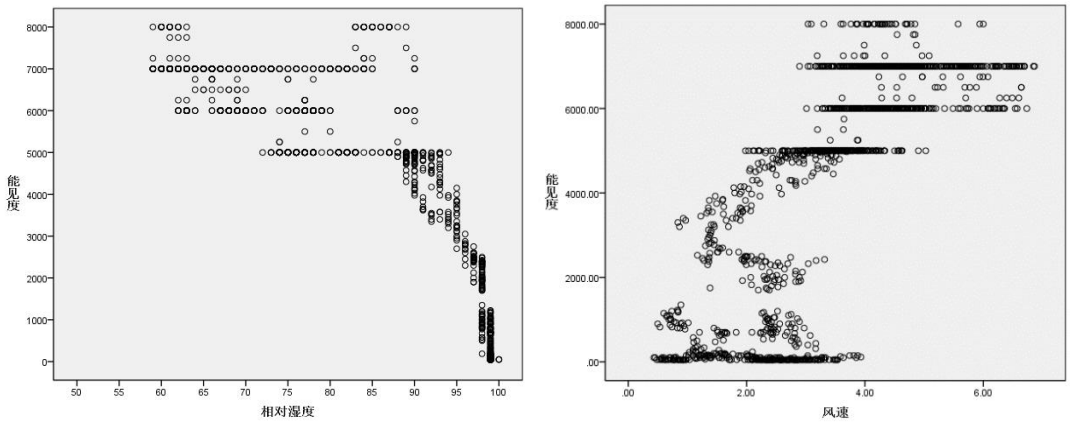
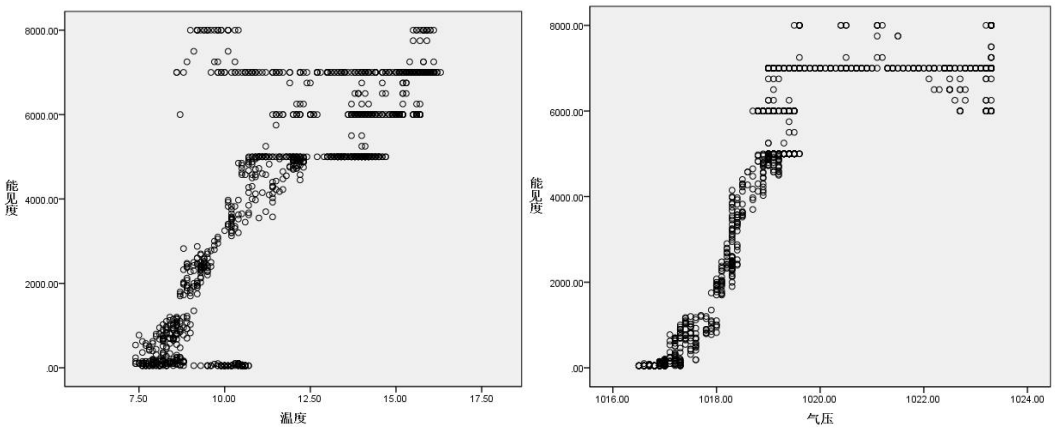


图 8 温度-能见度关系（左）和气压-能见度关系（右）



3.2 问题二模型的建立与求解

3.2.1 问题二的描述及分析

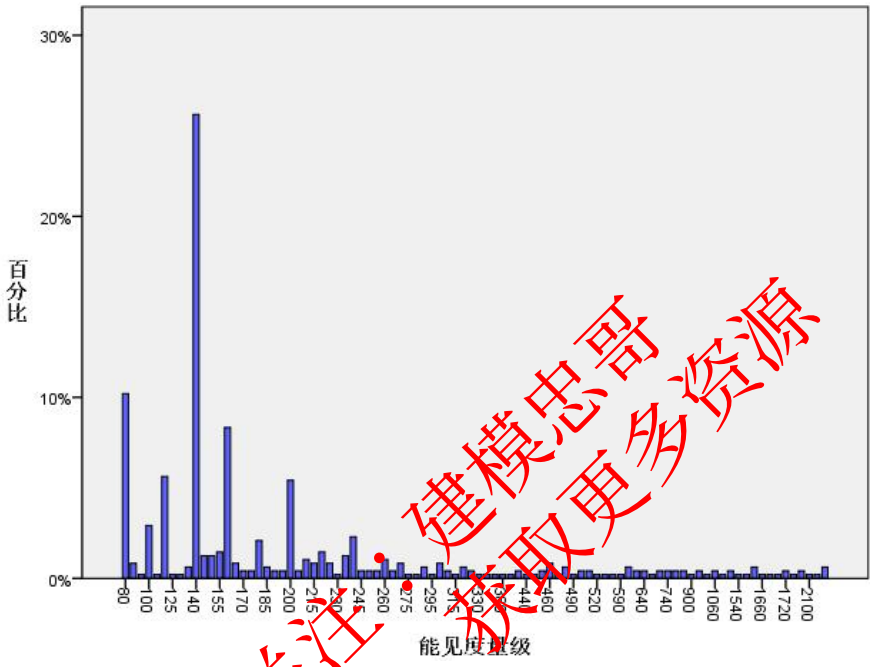
问题二想要建立一个基于视频数据的能见度估计深度学习模型，首先想到的是对视频数据进行处理，每隔 24 帧读取一帧图片作为当前时间的有效图片（总共得到了 40000 多张数据）。之后处理能见度数据表，在本项目中只关注于某几个重要的影响参量，将其余参量舍去之后得到一张包含所有重要参量的表。统计规律之后发现，仪器每分钟采集 4 个时间节点的能见度数据（某些分钟内采集 3 个时间节点的能见度数据），将数据进行归一化操作，4 条数据合成一个也即数据表中每分钟对应 1 条数据。此时能见度数据处理完毕，接着应调整有效图片数据以匹配能见度数据，每隔 60 张图片取一张有效图片作为本分钟的代表图。此时视频数据中生成的代表图的数据应与能见度表中对应时间的数据量应该一致。最终得到 463 个数据如下所示。

图 9 视频数据分类之后每分钟 1 张的数据

之后需要做的是处理能见度值数据，主导能见度变化的不仅有过去时次的能见度，还与风、温度、相对湿度等要素也有一定的关联[7]，我们重点关注的是具体的能见度数值，在表中能表现能见度值的主要有 RVR_1A 和 MOR_1A 这两个参量，查阅相关资料[8][9]。后得知在民用航空领域内使用最多的是 RVR 的概念，为了全面适应本题所给出的全部数据，本文决定对 RVR 和 MOR 的数据进行混叠处理，以经验值 RVR 比上 MOR 约等于 8 比 2，最终混合而成新的能见度数据。新数据如下表所示。在分析新生成的数据之后发现，能见度值处于 60 到 300 之间的数据最为多，顾将数据动态的分为以下情况，能见度值处于 0

到 300 之间的图片分割为 4 类，之后能见度值处于 301 到 500 的图片分割为 2 类，能见度值处于 501 到 750 的图片为一类，能见度值处于 751 到 1000 的图片为 1 类，能见度值大于 1000 的分为 1 类，最终得到 9 类分割后的数据。新生成数据的频次统计如下所示。新生成的数据如下所示。训练集目录结构如下所示。至此，本问题所需要准备的实验数据已经准备完毕。

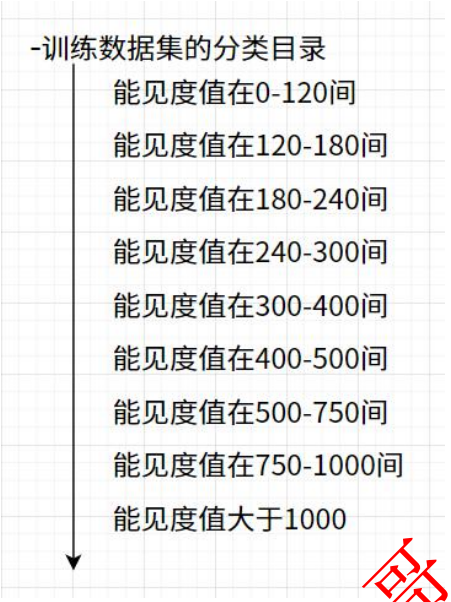
图 10 分类之后的各张图片能见度分布图



表格 7 各时间段混合能见度

北京时间	混合能见度（0.8*RVR_1A+0.2MOR_1A）
2020/3/13 0:02	2700
2020/3/13 0:03	2715
2020/3/13 0:04	2710
2020/3/13 0:05	2445
2020/3/13 0:06	2145
•••	•••
2020/3/13 7:56	90
2020/3/13 7:57	95
2020/3/13 7:58	110
2020/3/13 7:59	110

图 11 训练数据集目录分类



已经有国内外的研究学者对雾天能见度开展了诸多研究，Saadatseresht M 等利用人工神经网络（ANNs）对大气能见度预报进行了研究[10]；马楚旋等[11]利用遗传神经网络对能见度进行了预测；王继志等[12]对低能见度天气的预测方法做了研究。已经取得了有效的成功，基于前人研究成果，本文使用 CNN 网络来对数据进行处理。

之后将处理好的数据分为两部分，一部分作为训练数据集最终传入 CNN 中，另一部分作为验证网络训练结果的验证数据集。在本题目中，神经网络最终的输出为能见度值在以上 9 类中概率取值。

3.2.2 问题二模型的建立与求解

在本文中，为了提高 CNN 在雾霾识别的准确率，本文将数据集中所有图像块转换为 150*150（单位为 pixel）大小。一般来讲，数据集的制作需要大量的准确数据，但由于本题所给出的准确数据有限，有些分类的图片多有些分类的图片少，故本文通过数据集增广的方式，对图片数据集进行适当的扩充，考虑到短时间内难以获取到足够数量的雾霾图片，本文采用重复利用现有数据的方式来进行图像的增广，将图片数据集扩充到每一类有相同的数量，在本文中这个数量是 300 张。

卷积层（Conv）是 CNN 提取图像特征的关键。一般每个 Conv 都会有多个卷积核，每个卷积核的作用是用来提取一种图像特征。卷积计算表达式如下所示。

$$x_i^l = f(\sum_{i \in M_j} x_i^{l-1} * W_j^l + b_j^l) \tag{10}$$

其中 l 表示第 l 层 Conv，W_j^l 为该层的第 j 个卷积核的权值矩阵，*代表卷积运算，b_j^l 为该卷积核对应的偏置项，x_i^{l-1} 为网络上一层第 i 个输出灰度图，M_j 为输入层第 j 个卷积核的感受野。卷积运算过程中，需要设定单步卷积运算的步长，步长越大，提取特征越为稀疏，但步长过小，则可能导致特征提取过密从而产生冗余。

如果 CNN 仅仅只执行多层的卷积操作，那么实质上仅仅是对图像进行一种线形运算，单纯的增加卷积操作并不会对最后结果产生影响。因此，卷积运算结束后，通常采用激励函数，对卷积中的神经元增加非线性。本文使用的是四种常用激励函数中 ReLU 函数。函数表达式如下所示。

$$f(x) = \max(0, x) \tag{11}$$

Conv 之后，一般都会采用池化层（Pool）来池化 Conv 的输出，对模型特征进行降维，

降低计算机运算量，提升模型的训练效率，部分降低模型过拟合的可能性。Pool 计算表达式如下所示。

$$x_i^l = f(x_i^{l-1}) \tag{12}$$

其中，x_i^{l-1} 为当前输入的特征图层，x_i^l 为下采样得到的特征，$f()$ 表示池化的方法，一般采用平均池化（Mean pooling）和最大池化（Max pooling）两种方法。Mean pooling 是对邻域内的特征点进行平均，更多的保留图像的背景信息，Max pooling 计算邻域内的最大值，更好的保留图像的纹理信息。

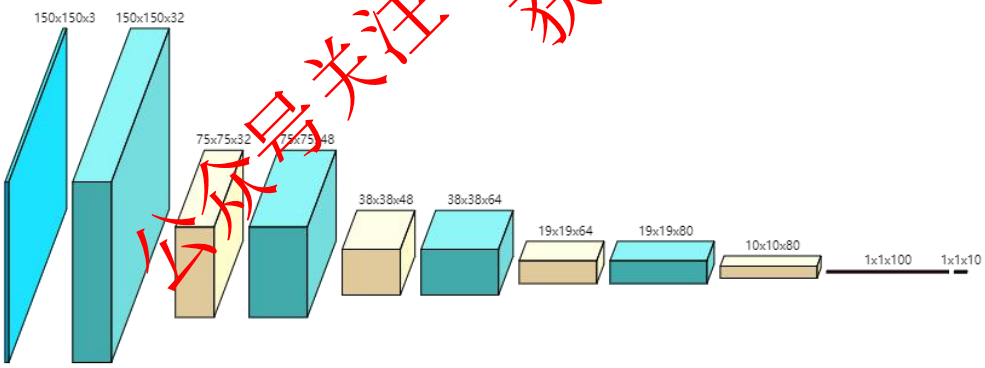
神经网络最后需要对输入数据进行分类，那么就希望网络在最后输出时能够输出输入图像所属类别的概率分布。因此，一般在 CNN 结构的最后一层也就是输出层之前加一个 Softmax 层。假定训练样本有 m 个，样本共有 n 个类别，其中 W 表示网络权重，$W_n^T x(i)$ 是 Softmax 层的输入，那么第 i 个样本属于第 j 类的概率可以表示为：

$$P(y^{(i)}=n \mid x^i;W) = \begin{bmatrix} P(y^{(i)}=1 \mid x^i;W) \\ P(y^{(i)}=2 \mid x^i;W) \\ \bullet\bullet\bullet \\ P(y^{(i)}=n \mid x^i;W) \end{bmatrix} = \frac{1}{\sum_{j=1}^{n} e^{W_j^T x(i)}} \begin{bmatrix} e^{W_1^T x(i)} \\ e^{W_2^T x(i)} \\ \bullet\bullet\bullet \\ e^{W_n^T x(i)} \end{bmatrix} \tag{13}$$

其概率分布之和为 1[13]。

最终，经过以上的分析，最终确定了模型的配置，模型结构如下图所示。输入是 150*150（单位为 pixel）的灰度图像，输出为 2 维向量。下表中给出了模型各层的具体参数。

图 12 神经网络模型各层之间的具体参数



表格 8 CNN 网络各层具体设计

层类型	卷积核大小	输出大小	步长
Conv1	5*5*32	150*150*32	1
Pool1	3*3*32	75*75*32	2
Conv2	3*3*48	75*75*48	1
Pool2	3*3*48	38*38*48	2
Conv3	3*3*64	38*38*64	1
Pool3	3*3*64	19*19*64	2
Conv4	3*3*80	19*19*80	1
Pool4	3*3*80	10*10*80	2
FC1	10*10*80	1*1*100	
FC2	1*1*100	1*1*9	

使用上述设计的网络对 3.2.1 中所设计的数据集进行求解，最终推导得到一个用于对雾霾图像分类的能见度检测模型。图像训练过程如下图所示。能见度检测过程如下图所示。

图 13 神经网络模型训练构成

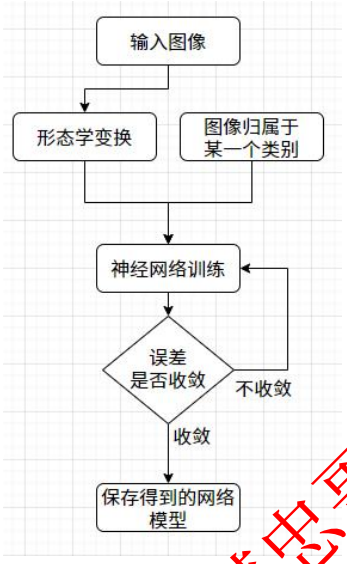
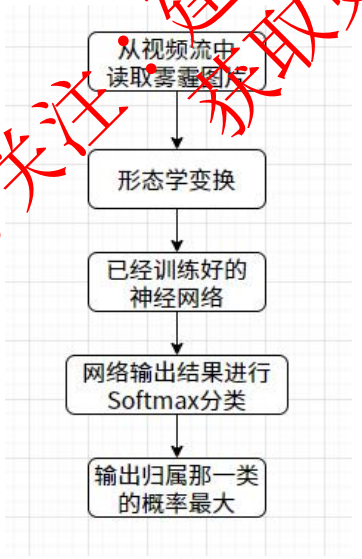


图 14 神经网络预测过程



使用上文留下的测试数据集来对网络进行测试，实现了 78.8%的识别成功率。训练集的测试情况如下表所示。

表格 9 CNN 模型在测试集上的正确率

计数	图像原地址	程序输出内容	程序输出分类	原始分类
1	./data/2/pic_096.jpg	能见度在 180-240 之间	2	2
2	./data/5/pic_024.jpg	能见度在 400-500 之间	5	5
3	./data/4/pic_208.jpg	能见度在 240-300 之间	4	3
4	./data/3/pic_223.jpg	能见度在 240-300 之间	3	3
5	./data/3/pic_287.jpg	能见度在 240-300 之间	3	3
	•••	•••		
	./data/6/pic_248.jpg	能见度在 500-750 之间	6	6
	./data/6/pic_102.jpg	能见度在 500-750 之间	6	6
	./data/4/pic_255.jpg	能见度在 400-500 之间	4	5
500	./data/0/pic_044.jpg	能见度在 120-180 之间	0	1
正确率				77.75%

之后再对机场视频数据剩下的图片应用网络模型来进行能见度值的预测，预测数据如下表所示，预测数据每分钟截取四张代表图片进行判断，下表仅显示每分钟一张。

表格 10 CNN 模型在真实数据上的输出

计数	时间	程序输出内容	程序输出分类
1	08:00	能见度在 120-180 之间	1
2	08:01	能见度在 120-180 之间	1
3	08:02	能见度在 120-180 之间	1
4	08:03	能见度在 120-180 之间	1
5	08:04	能见度在 0-120 之间	0
	•••	•••	
905	11:43	能见度在 400-500 之间	5
906	11:44	能见度在 400-500 之间	5
907	11:45	能见度在 400-500 之间	5
908	11:46	能见度在 400-500 之间	5

3.3 问题三与四模型的建立与求解

3.3.1 问题三与四的描述及分析

能见度可表示为将目标物从周围环境中识别出来的距离，因此能见度就与目标物与背景的亮度对比度有关。目标物越暗背景越亮或目标物越亮背景越暗则对比度越大，能见度也越高。要想从大雾天气图像中获取能见度信息，就需要用到大气散射模型，将能见度的求取问题转化为求取大气消光系数的问题。利用暗通道先验理论对图片进行暗通道处理，提高目标识别点的对比度，从而通过对目标点大气透射率的测量间接得到图像的能见度信息。

测量特定目标点的透射率需要知道目标点到传感器的距离，而对于视频监控图像，成像过程中缺失了深度信息，从而目标距离需要通过几何法来测量。几何法利用透视投影的交比不变性，通过找出图像中的灭点和灭线经过计算得到结果。

3.3.2 问题三与四模型的建立

Koschmieder 通过对大气中亮度的衰减进行研究[14]，提出了有关大气消光系数的关系式。

$$L = L_0 e^{-\beta d} + L_\infty (1 - e^{-\beta d}) \tag{14}$$

d 为目标物的距离，L 为目标物的固有亮度，L_0 为目标物周围环境的亮度，L_∞ 为天空大气亮度。Duntley 在上式的基础上将亮度模型转化为亮度对比度模型[15]，提出对比度衰减定律：在距离 d 处，物体相对于背景的对比度为 C_0 ，那么物体只能在对比度为C时才能被人眼感知。

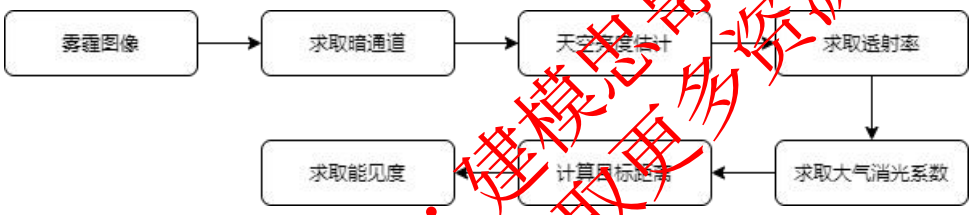
$$C = C_0 e^{-\beta d} \tag{15}$$

根据该式便可计算出能见度 V。根据国际照明委员会(CIE)给出的定义：在对比度阈值为 0.05 时，人眼所能观测到物体的最大距离称为气象能见度距离。在目标物相对于周围环境的对比度 C_0=1 时，就可以推导出能见度的一般公式[16]。

$$V_{mvd} = -\frac{\ln 0.05}{\beta} \approx \frac{3}{\beta} \tag{16}$$

可以看出，能见度的测量可以转化为测量大气消光系数来解决。

图 15 基于暗通道先验的能见度估计算法流程



3.3.3 问题三与四模型的求解

模型三与四的求解主要分为两个部分，以下分别对两个部分进行阐述。

第一部分是，基于暗通道先验的大气透射率获取。在计算机视觉和数字图像处理领域，一般用以下大气散射模型来描述雾霾图像。

$$I(x) = J(x)t(x) + A(1 - t(x)) \tag{17}$$

其中，x 示数字图像中的像素点坐标，I(x)是观测到的图像，J(x)是目标物场景强度，A 是大气环境光线（主要用图像中最亮的 0.1%点的均值代替），t(x)是大气透射率，描述了目标物经媒介传播后未到达摄像头的光线所占比率。

暗通道先验理论的内容是：针对室外无雾霾彩色图像任意区域（天空区域除外），R、G、B 三通道中，总是存在亮度最小的像素值[17]。即该图像中以任一像素为中心的窗口区域中，存在某一通道的光强最小值接近于 0。这个统计规律就是暗通道先验算法的主要思想，具体的数学定义如下。

$$J^{dark}(x) = \min_{c\in\{r,g,b\}} \left(\min_{y\in\omega(x)} J^c(y) \right) \tag{18}$$

其中，J^c是 RGB 图像的某个通道的光强值；ω(x)代表以 x 像素点为中心的一个窗口，目前选择经验值 15*15。对式(4)做暗通道处理得得到以下结果。

$$\min_{c\in\{r,g,b\}} \left(\min_{y\in\omega(x)} I^c(y) \right) = t(x) \min_{c\in\{r,g,b\}} \left(\min_{y\in\omega(x)} J^c(y) \right) + (1 - t(x))A^c \tag{19}$$

根据暗通道先验理论可以得知，在一幅图像中，除了天空区域，其他像素点的J^{dark}总是趋向于 0，即有以下定义。

$$J^{dark}(x) = \min_{c\in\{r,g,b\}} \left(\min_{y\in\omega(x)} J^c(y) \right) \approx 0 \tag{20}$$

由以上公式可得，大气透射率公式为。

$$t(x) = 1 - \min_{c\in\{r,g,b\}}\left(\min_{y\in\omega(x)}\left(\frac{I^c(y)}{A^c}\right)\right) \quad (21)$$

由于在晴朗天气时，摄像机拍摄图像仍会受到少量雾霾的影响，因此需引入权重因子 ω，将上式矫正后得到（ω值越大，去雾效果越好，此处我们选择 0.85）。

$$t(x) = 1 - \omega \min_{c\in\{r,g,b\}}\left(\min_{y\in\omega(x)}\left(\frac{I^c(y)}{A^c}\right)\right) \quad (22)$$

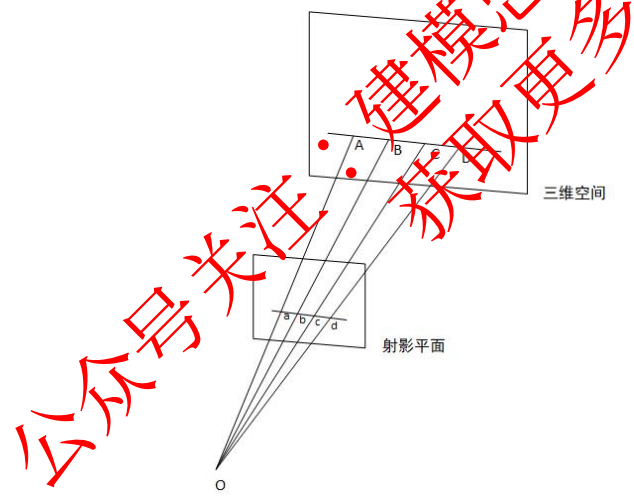
推导得到大气消光系数 β 的计算公式为。

$$\beta = \frac{\ln(1/t)}{d} \quad (23)$$

第二部分是，目标物与摄像机之间距离 d 的获取：透视投影法。首先利用交比不变性由射影几何可知，三维空间中的平行直线投影到特定的射影平面上时会相交于一点，称为射影平面中的灭点。一组三维空间中的平行直线可以确定一个灭点，一个平面上所有平行线组确定的灭点共线，称为灭线。

根据透视投影和射影变换原理，图像中共线点的交比与真实世界中的交比相等，据此可计算该问题中目标点的距离。

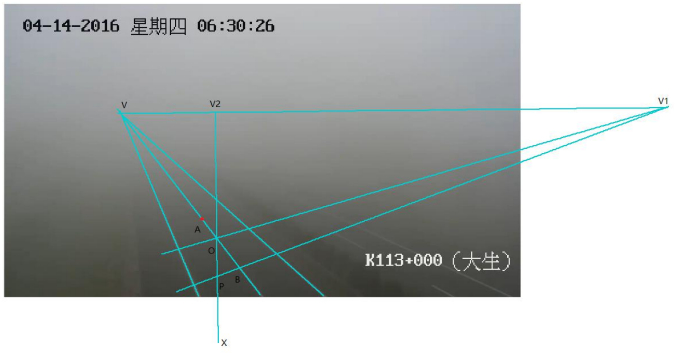
图 16 射影变换示例图



$$\mathrm{CR} = \frac{|A-C||B-D|}{|A-B||C-D|} = \frac{BD}{CD} \qquad \mathrm{CR} = \frac{|a-c||b-d|}{|a-b||c-d|} = \frac{bd}{cd}$$

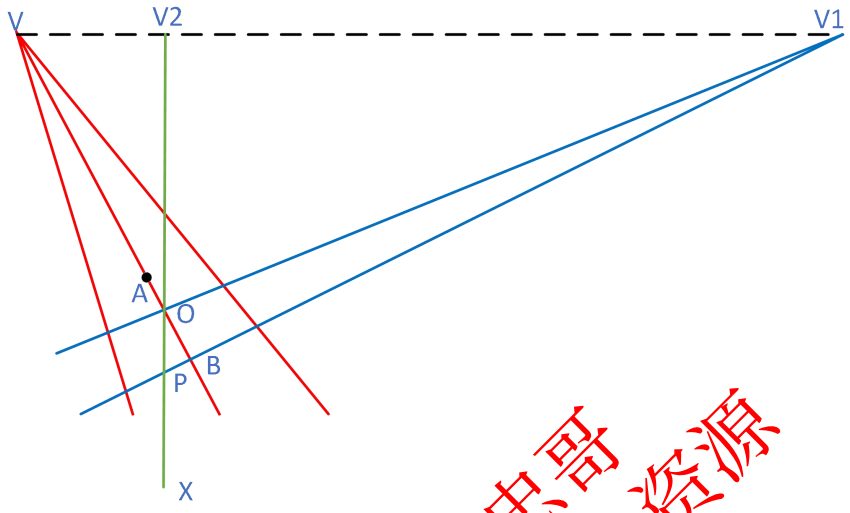
然后再去求取目标物距 d。

图 17 射影变换示意图



根据高速公路道路标线的平行性和左右车道中分界线（虚线）的平行性，找到了两组灭点 V、V1，并构建出灭线（$\overline{VV1}$）。经简化后如下图所示。

图 18 图 17 简化后的模型



点 A 为左侧车道中间虚线的上端点，点 O 为左侧车道中间虚线的下端点，点 B 为左侧车道靠下虚线的上端点。已知 AO 间距离为 6m、OB 间距离为 9m。自灭线上取一点 V2，连接 V2 与 O，与蓝色线相交于 P 点，直线 V2O 近似垂直于灭线。设点 X 为摄像机到地面投影位置。图中点的像素坐标为。

表格 11 图像中点的坐标

点	V	A	O	B	P	V2
坐标	(257.459,281.992)	(520,484)	(575.116,522.98)	(649,585)	(671,525)	(277.032,521.758)

图中点 V、A、O、B 四点共线，根据交比不变性，有如下的公式。

$$\mathrm{CR} = \frac{\|A-B\|\|V-O\|}{\|O-B\|\|V-A\|}$$

得到交比为 1.6741，经验证与 $\frac{AB}{AO} = \frac{15}{9} = 1.6667$相符。

同理，点 V2、O、P、X 四点也共线，利用交比不变性原理，有如下公式。

$$\mathrm{CR} = \frac{\|X-O\|\|P-V2\|}{\|X-P\|\|O-V2\|} = \frac{15}{9}$$

由上式可求得目标点 O 到摄像机的水平距离||X-O||为 459.384。

点 O 到点 V2 的距离与点 O 到摄像机的高度相等，已知摄像机高度为 6m，则点 O 到到点 X 的距离为 9.25m，根据勾股定理求得点 O 到摄像机的实际距离 d 约为 11m。

第三部分，能见度求取。我们以左侧道路中间虚线的下端点 O 为目标识别点，通过计算每张图片中该点的大气透射率，代入公式从而得到一组能见度值，对应每张图片的能见度。该时间段能见度随时间变化曲线为。

图 19 原图（左）、透射率图（右）去雾后图（下）

图 20 能见度随时间变化曲线图

由能见度曲线可以知道，能见度在短时间内有小幅波动，总体呈现缓慢上升趋势。在高速公路实际路况中，由于风的因素，摄像机观测到的能见度会有小幅变化；由于图片时间是在早晨，随着时间推移，白天的天气亮度逐渐增强，温度升高，雾霾有逐渐退散趋势，因此能见度逐渐升高，与实际情况相符。

对问题三得到的能见度数据进行多项式拟合，多项式的一般形式为。

$$y = p_0 x^n + p_1 x^{n-1} + p_2 x^{n-2} + p_3 x^{n-3} + \cdots + p_n \quad (24)$$

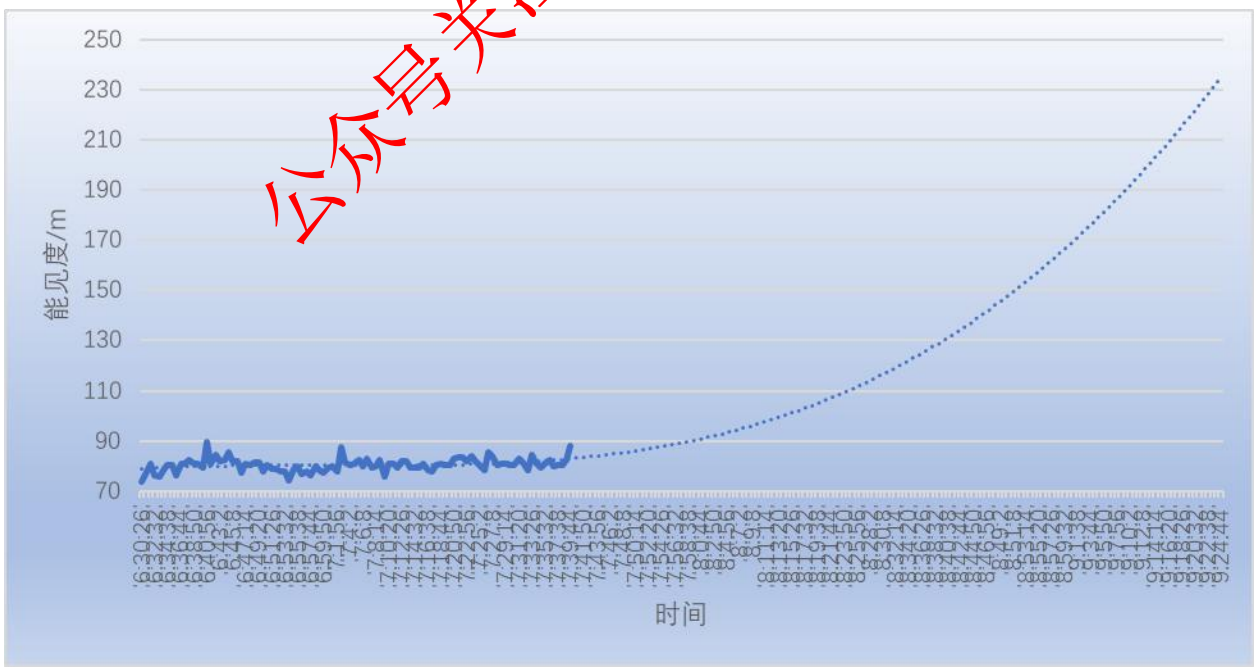
多项式拟合的目的是为了找到一组系数 p_0，p_0，…，p_n，使得拟合方程尽可能的与实际样本数据相符合。结合实际能见度变化趋势和拟合误差最小原则，我们得出能见度随时间的变化关系式为。

$$V = 2 \times 10^{-5}x^3 - 0.0025x^2 + 0.1086x^1 + 78.73 \quad (25)$$

图 21 拟合能见度变化曲线图

x 为 1 到 100 的整数，代表从时间 2016-4-14 日 6:30:26 起每隔 42s 的时间戳。由预测趋势图可以看出，能见度随着时间的增长逐渐增大，雾逐渐消散。到 8:50 左右能见度增长到 150 米，到 11:00 左右能见度达到 1000 米。

图 22 能见度预测图



4 模型的结论与评价

4.1 模型一的结论与评价

通过对问题一建立模型求解能见度与气象因子之间的关系，建立回归方程，得到模型结果图，可以得出以下结论：

能见度是各类气象因子共同作用结果，其中风速、温度、相对湿度、气压对能见度的影响较大，且在因变量与自变量相关分析中可知，相对湿度对能见度呈高度负相关，温度、气压、风速呈正相关，这与文献[4]结果一致。

在两组数据的回归模型中，能大致反应当日的能见度变化趋势，但模型存在一定不足：在低能见度区时能见度动态变化比较平稳，模型往往偏离观测值较大，甚至出现负能见度的错误预测。在高能见度区，观测值得动态变化往往很大，模型动态变化则相对较小。

气象因子并不是影响能见度的唯一因素，空气中污染物浓度也会对能见度产生影响，各个气象因子在不同季节也有不同的变化。因此两批数据的模型并不完全相同，也说明预测模型仅能做短时预测，无法找到能见度与气象因子之间的经验公式。

以上提出能见度与气象因子的回归模型，但由于仅有两批数据做模拟，模拟结果并不是很准确，因此上述的模型方法仍可以优化得到最优解。比如，在相关性分析中我们假定自变量与因变量存在线性或能通过简单变换得到线性的非线性，所没有考虑复杂的计算模型，所以得到的回归方程十分复杂。再比如，我们仅使用 MOR_1A、WS2A 等数据，一分钟内各气象因子的变化数据，没有参考长时间范围的能见度、风速等变化，同时做回归时仅考虑 MOR（气象光学视程）能见度，没有考虑 RVR（跑道视程），这都会对结果产生一定影响，多元回归分析优化问题方法较多，常见 BGD、SGD 与 Mini-batch SGD 等多类优化算法，但由于算力有限，无法对问题进一步优化处理。

4.2 模型二的结论与评价

由于导致能见度发生变化的因素比较多，变量系统较为复杂，尤其是在处理低能见度天气时，风速、湿度等因素发生微小变化，也会导致最后神经网络训练难以收敛。

在模型二的处理中，采用了较为模糊的分类 CNN，最终的输出结果是能见度值在某一区间的可能性最为大，不能够对能见度输出一个精确的数值。但是本文在分析现有数据的情况下，对分类做了动态的划分，将能见度较为低的情况划分了足够的类别，最终在有限条件下取得了 77.8%的模型识别率。

后期如果要优化算法，只要应该从分类的细致程度上入手，或者考虑不使用分类的方法，而是使用每一张图对应着一个标签，最终通过均方误差函数最终实现能够对一副图片输出精确的能见度值。

4.3 模型三与四的结论与评价

针对来自视频的单幅图像的信息分析，本文应用了大气散射模型和暗通道先验算法对图片中的大气能见度进行估计，得到了图像的透射率图和去雾后的图像。通过查找资料，我们挖掘出图像中隐藏的部分尺寸信息，利用几何数据的约束性进行建模，从而得到能见度的估计的支撑数据。通过对问题三、问题四求解出的能见度估计和预测进行分析，发现估计的能见度值在实际允许的可靠范围内，预测的能见度变化趋势与实际变化趋势相符。但由于模型的固有特性，只利用了特定目标点进行能见度估计，所得到的的估计值可能不够准确，带来较大误差。题目给出的信息为连续视频信息中截取出的图片信息，由于天气因素，相邻估计值变化较大，也会影响预测值的准确性。

参考文献

[1] 许艳丽,李海波,成孝刚,邵文泽,吕泓君. 一种基于深度学习的雾霾能见度检测方法[P]. 江苏：CN107274383A,2017-10-20.

[2] 凌强,陈春霖,李峰. 一种基于深度学习的能见度检测方法[P]. 安徽省：CN107506729B,2020-04-03.

[3] 郄岩. 雾天能见度检测与预测方法研究[D].河北科技大学,2019.

[4] 王淑英,张小玲,徐晓峰.北京地区大气能见度变化规律及影响因子统计分析[J].气象科技,2003(02):109-114.

[5] 华丽君,王剑峰,杨仲玮.兰州新区 2017—2019 年环境空气质量变化趋势研究[J].环境研究与监测,2020,33(02):66-71.

[6] 杨树成. 应用统计学[M]：成都：西南交通大学出版社，2017.05,192-201.

[7] 朱国栋,杜安妮. 深度学习在机场能见度预测中的应用[A]. 中国气象学会.第 34 届中国气象学会年会 S20 气象数据：深度应用和标准化论文集[C].中国气象学会:中国气象学会,2017:7.

[8] 杨瑜, 丁文敏. 浦东机场低跑道视程变化特征及其影响机制分析[J]. 干旱气象, 2016, 34(005):873-880.

[9] 王勇.民航地面气象观测中的能见度分析[J].科技经济导刊,2020,28(03):72+71.

[10] Saadatseresht M , Varshosaz M . Visibility prediction based on artificial neural networks used in automatic network design[J]. The Photogrammetric Record, 2007, 22(120):336-355.

[11] 马楚焱,祖建,付清盼,罗凌霄.基于遗传神经网络模型的空气能见度预测[J].环境工程学报,2015,9(04):1905-1910.

[12] 王继志,杨元琴,周春红,等.雾霾低能见度天气分析与预测方法研究[C].//中国气象学会.2007 年中国气象学会年会论文集.2007:145-149.

[13] 宋文豪. 基于机器视觉的核燃料芯块表面裂纹检测方法研究[D].郑州大学,2019.

[14] W.E.K. Middleton. Vision through the atmosphere [M] // Geophysik II/Geophysics II. Berlin: Springer, 1957:254 —287.

[15] C STEFFENS. Measurement of visibility by photo-graphicphotometry [J]. Industrial & Engineering Chemistry, 1949, 41(11):2396 —2399.

[16] 许艳丽.雾霾天气条件下能见度的检测与恢复算法研究[D].江苏:南京邮电大学,2018.

[17] He K, Sun J, Tang X. Single image haze removal using dark channel prior[C]//Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on. IEEE, 2009:1956-1963.