

# 基于FCM聚类的形势场图像分割方法研究

胡光亮, 段 勇

沈阳工业大学, 信息科学与工程学院, 辽宁 沈阳

收稿日期: 2022年3月4日; 录用日期: 2022年3月17日; 发布日期: 2022年4月1日

## 摘 要

本文研究基于FCM聚类算法的天气形势场图像分割问题, 首先通过解析将气象中心历史再分析资料中多种形势场数据可视化绘制出高空天气图像, 再结合本文所研究的问题对FCM聚类算法的原理以及步骤进行描述, 然后将FCM聚类算法应用于高空天气图像以及形势场数据可视化过程中可能带有噪声的高空天气图像的分割, 最后通过实验结果验证FCM聚类算法能够较好地实现形势场可视化图像的分割, 进而标注出高空天气图像中等值线。

## 关键词

图像分割, FCM聚类, 高空天气图, 形势场

# Research on Situation Field Image Segmentation Based on FCM Clustering

Guangliang Hu, Yong Duan

School of Information Science and Engineering, Shenyang University of Technology, Shenyang Liaoning

Received: Mar. 4<sup>th</sup>, 2022; accepted: Mar. 17<sup>th</sup>, 2022; published: Apr. 1<sup>st</sup>, 2022

## Abstract

In this paper, the FCM clustering algorithm is used to solve the problem of weather situation field image segmentation. First, the high-altitude weather images are drawn by analyzing and visualizing various situation field data in the historical reanalysis data of the meteorological center. Then, the principle and steps of FCM clustering algorithm are described based on the problems studied in this paper. Then the FCM clustering algorithm is applied to the segmentation of the high-altitude weather images and the high-altitude weather images that may contain noise in the process of situation field data visualization. The experimental results show that the FCM clustering algorithm can achieve better segmentation of the situation field visualization images, and then the isolines

in the high-altitude weather images can be marked.

## Keywords

Image Segmentation, FCM Clustering, High Altitude Weather Map, Situation Field

Copyright © 2022 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

气象领域中形势场数据包括位势高度场、温度场、湿度场、比湿度场等, 它们经常参与在各种天气相似预测。而天气图是形势场数据的可视化图像, 包含位势高度、温度、湿度、风速、风向等信息, 它反映了一定时间和一定区域的天气情况, 并结合图上的高度绘制等高线, 从而显示空间上的天气情况和天气系统的分布。通过对地面天气图和高空天气图的三维分析, 可以分析未来的天气变化, 这是目前传统天气预报的主要依据之一, 其中高空天气图占主导地位。因此, 如果可以对可视化后的形势场数据即高空天气图, 从中由计算机自动进行等值线特征的提取和对别分析, 将有效提高天气预报的效率和可能性。

为了解决上述问题, 通过借鉴图像分割的思想实现形势场数据的区域分割, 进而进行等值线的提取以及天气现象的分析和预测。

在学者们提出的图像分割的解决方法中, 基于聚类的图像分割是最常用的方法之一[1] [2]。陈建文等人[3]提出了基于 K-Means 的时间季节反演识别方法, 利用 HSV 图像色彩模型, 对季节图片的特征进行分析和研究; 但由于图像区域边界类别很难由 K-means 聚类算法在实际应用中准确区分, 基于此模糊 C 均值聚类(FCM)对 K-means 聚类算法做出改进, 提出了隶属度概念, 利用隶属度来判断每个类别的隶属度。陈建文等人[3]将图像聚类算法用在岩石图像分析, 为岩石识别和分类奠定了基础; Ngo L T 等人[4]提出了改进的 FCM 算法, 解决了多光谱卫星图像的多时间变化检测问题; 胡学刚等人[5] [6]利用非局部空间信息构建和图像, 实现彩色图像在均值聚类过程中, 聚类中心的自动选择; 冯仁光等人[7]提出改进的自适应权重聚类结合小波阈值去噪, 解决了矿物浮选过程中图像灰度分布特殊性难适应问题; Feng L [8]通过模糊均值聚类算法利用基于 Harris 角点检测的凸包理论, 提高了算法分割精度并缩短了计算时间; Wu S [9]等人通过将距离与密度相结合的距离算法, 克服了聚类中心选择不稳定缺点; 沈晓[10], Huang H [11]将脑部核磁和 FCM 相结合, 解决了脑部图像灰度不均匀以及噪声等问题。

基于此, 本文研究了历史形势场数据的解析与可视化处理方法以及模糊 C 均值聚类算法在天气图像分割中的应用。历史形势场数据在解析为高空天气图过程中会受到噪声和区域不均匀性的影响而导致等值线边界误分类的现象, FCM 聚类算法通过考虑计算隶属度值来对相邻像素的精确的分类实现图的精确分割。

## 2. 形势场数据及可视化处理

### 2.1. 建立形势场数据集

本文建立的用于分析的形势场数据集包括三种天气资料: 第一种是从欧洲气象中心获取的近 40 年内

的历史再分析资料, 具体包括 1980 年~2019 年的多个等压面上的位势高度场、温度场、气压场、比湿场等历史形势场数据; 第二种是第一种相对应的近 40 年的历史实况资料, 具体包括 1980~2019 年的降水量、能见度、风向风速等历史实况数据; 第三种资料是从辽宁省气象站实时获取的 EC 数值预报资料, 例如 2020 年 10 月 30 日的温度场、气压场等预报场数据。如图 1 所示。

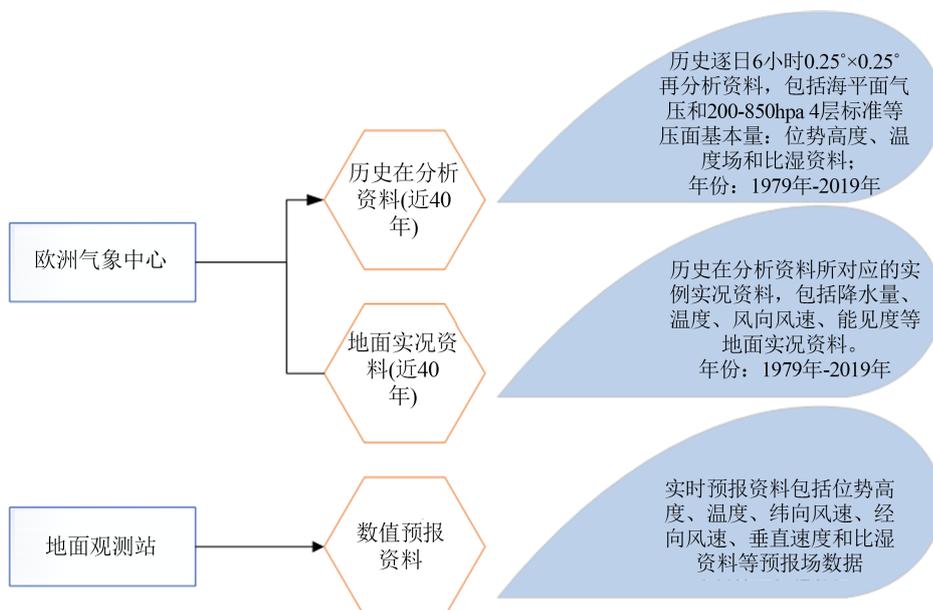


Figure 1. Source of data set of the situation field  
图 1. 形势场数据集聚源

分析资料包括高度场、温度场、湿度场和海平面气压场。这些历史再分析资料中包含多种属性, 包括经度、纬度、时次、层次, 经度范围在东经 60~150 度, 纬度范围在北纬 0~60 度, 每个文件为一天内的数据, 包含逐 3 小时的 8 个时次, 层次分为 200、500、700、850 共 4 层, 每种再分析资料的主属性不同, 主属性分别包含高度、温度、比湿等能够代表天气情况的气象要素。如图 2 所示。

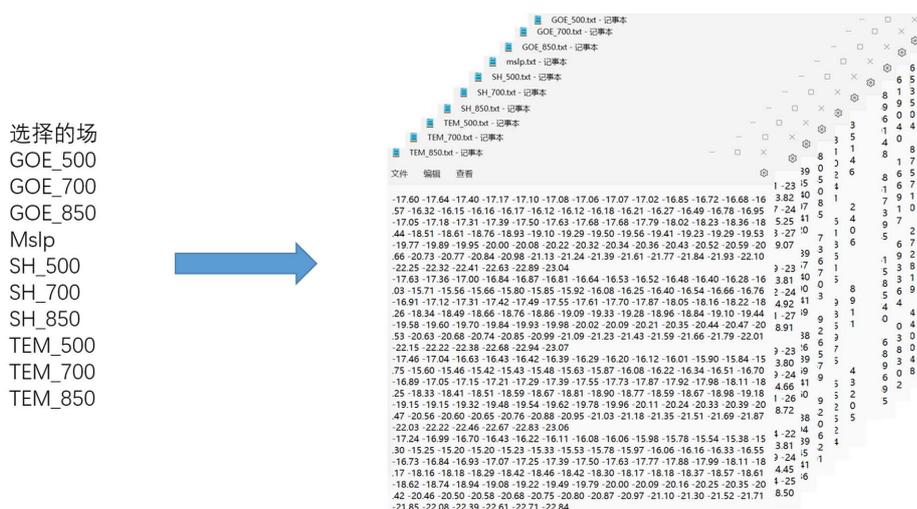


Figure 2. Data of situation field  
图 2. 形势场数据

## 2.2. 形势场数据可视化

根据本文的需要将形势场数据中的 500 hpa 高度场数据进行可视化设计, 对区域划分利用 Basemap 进行地图叠加将其解析为高空天气图像。如图 3 所示。

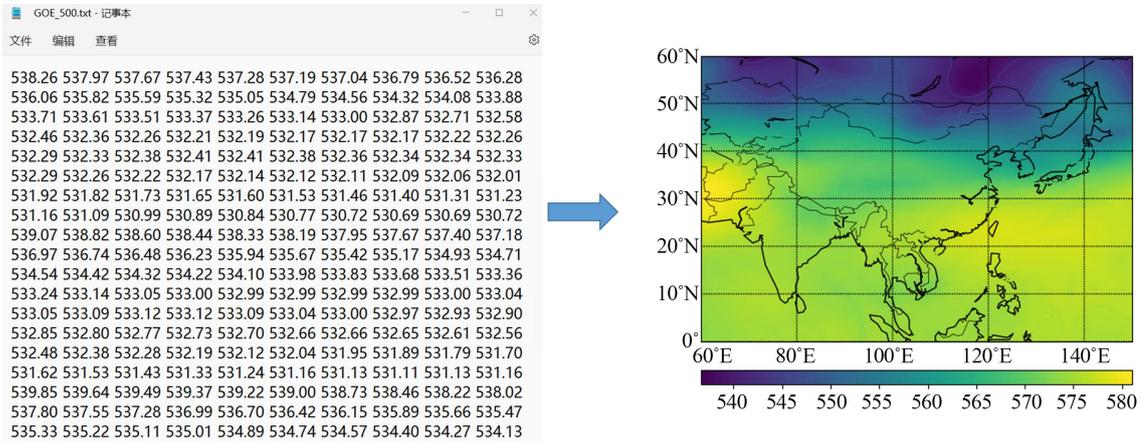


Figure 3. The altitude field data from the historical reanalysis data are analyzed into the upper-air weather image  
图 3. 历史再分析数据中高度场数据解析为高空天气图像

根据需要分析分区域的经纬度范围, 将解析的图像区域缩小以实现区域的裁剪。裁剪后高空天气图像如图 4 所示。

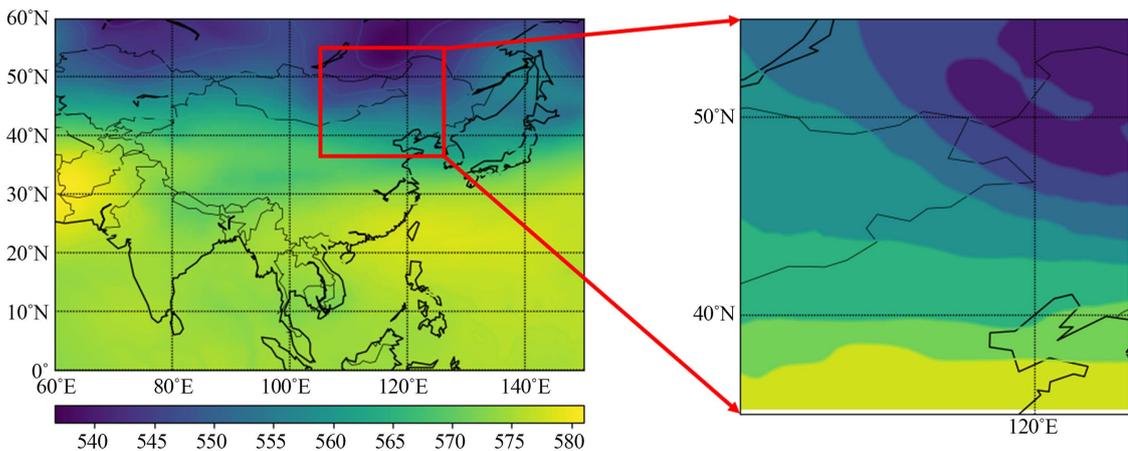


Figure 4. A clipped image of the weather aloft  
图 4. 裁剪之后的高空天气图像

此外, 可以将形势场数据进行三维可视化, 例如图 5 为 500 hpa 高度场立体结构。500 hpa 等压面的高度值, 随着高度增加, 气压值降低, 由于天气系统的影响, 降低程度不同, 颜色越深的区域单位面积气压值越小, 由此可以通过分析高空天气图像和高度场立体结构得出不同区域出现某种天气现象的概率。

## 3. 模糊 C 均值聚类应用于高空天气图像分割

### 3.1. 模糊 C 均值聚类(FCM)算法原理

模糊聚类[12][13][14]是一种将模糊数学理论及思想应用于非监督机器学习的方法, 通过模型数学的

隶属度等概念来确定待聚类样本属于某聚类类别的可能性。在许多情况下, 类的界限并不是绝对的。与  $k$  均值聚类相比, 模糊  $C$  均值聚类能得到更灵活的结果。因为在大多数情况下, 数据集中的对象不能划分为特定的集群, 所以将对象分配给特定的集群有点生硬, 而且还可能发生错误。所以, 对每个样本和各聚类簇赋予一个隶属度值, 该值表示该样本对象数据这个聚类的可能程度。假定样本集为  $X = \{x_1, x_2, \dots, x_n\}$ , 聚类类别为  $C$  个簇, 第  $i$  簇的聚类中心用  $C_i$  表示, 通过最小化目标函数(1)和约束条件(2)来实现聚类。公式如下:

$$J = \sum_{i=1}^c \sum_{j=1}^n u_{ij}^m \|x_j - C_i\|^2 \quad (1)$$

$$\sum_{i=1}^c u_{ij} = 1, j = 1, 2, 3, \dots, n; u_{ij} \in [0, 1] \quad (2)$$

其中  $c$  表示聚类中心的个数,  $n$  为样本的个数,  $m$  是一个隶属度因子。  $u_{ij}$  表示样本  $j$  对于类  $i$  的隶属度。

基于拉格朗日乘法将公式(1)转换如下的公式(3)。

$$J = \sum_{i=1}^c \sum_{j=1}^n u_{ij}^m \|x_j - C_i\|^2 + \lambda_1 \left( \sum_{i=1}^c u_{i1} - 1 \right) + \dots + \lambda_j \left( \sum_{i=1}^c u_{ij} - 1 \right) + \dots + \lambda_n \left( \sum_{i=1}^c u_{in} - 1 \right) \quad (3)$$

计算出隶属度函数和聚类中心  $C_i$  的迭代表达式(4)和(5)。

$$u_{ij} = \frac{1}{\sum_{k=1}^c \left( \frac{\|x_j - C_i\|}{\|x_j - C_k\|} \right)^{\frac{2}{m-1}}} \quad (4)$$

$$C_i = \frac{\sum_{j=1}^n (u_{ij}^m x_j)}{\sum_{j=1}^n u_{ij}^m} \quad (5)$$

最终 FCM 算法通过重复迭代公式(4)的隶属度和公式(5)的簇聚类中心计算, 得到约束目标函数的极值, 从而完成聚类过程。

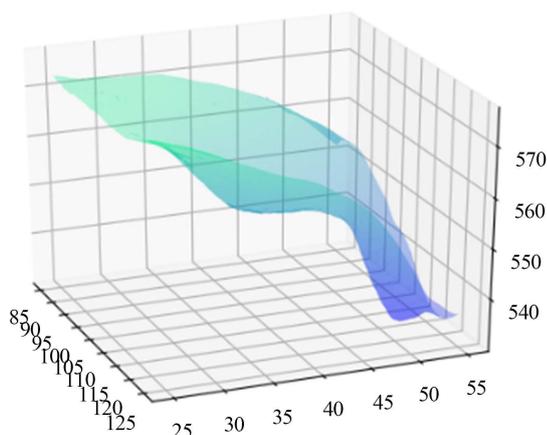


Figure 5. 500 hPa height field three-dimensional structure  
图 5. 500 hPa 高度场立体结构

### 3.2. 模拟添加噪声

由于形势场数据中存在数据缺失情况, 在数据可视化过程中可能产生噪点[15], 扰乱图像的可观测信息, 使图像不清晰, 为观察噪声图像对本文实验的影响, 通过添加噪声来模拟生成常见的高斯噪声和椒

盐噪声图像。

由形式场数据可视化的天气图原始图像如图 6 所示。高斯噪声是指概率密度函数服从高斯分布(即正态分布)的一类噪声, 通过函数设定噪声均值和方差, 得到其噪声图像如图 7 所示; 椒盐噪声又称为脉冲噪声, 它随机改变一些像素值, 是由图像传感器、传输信道、解码处理等产生的黑白相间的亮暗点噪声。通过函数设定其随机噪声比例得到噪声图像如图 8 所示。



Figure 6. Original upper-air weather image  
图 6. 原高空天气图像

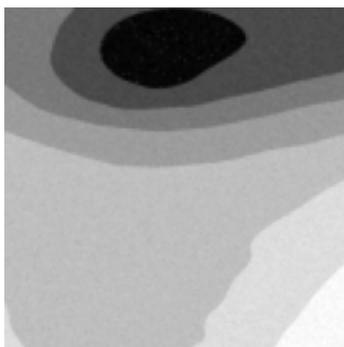


Figure 7. Gaussian noise image  
图 7. 高斯噪声图像

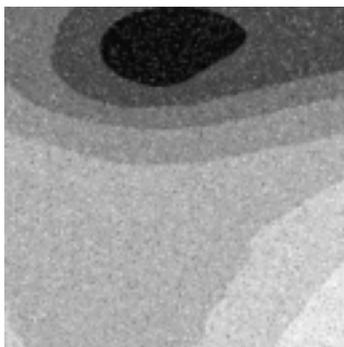


Figure 8. Salt and pepper noise image  
图 8. 椒盐噪声图像

### 3.3. 基于模糊 C 均值聚类算法的天气图像分割步骤

首先将高空天气图裁剪为本文研究的经纬度范围, 然后初始化高空天气图的聚类中心、隶属度矩阵、

迭代次数, 确定聚类的类别数目和迭代的次数或者两次相邻聚类中心的差值, 指定模糊加权指数值。根据公式(2)的约束, 定义隶属度矩阵每列的和为 1, 利用公式(4)进行隶属度计算并更新隶属度矩阵。根据公式(5), 利用隶属度矩阵计算高空天气图的聚类中心。在每次迭代中通过公式(3)获得目标函数的数值, 并和设定的阈值(缺省值为  $1.0e-6$ )进行比较, 从而确定是否满足迭代停止条件。当满足目标函数值小于缺省值条件时, 输出聚类结果和聚类簇中心, 生成聚类图像; 如果不满足条件则继续判断实验是否已经达到程序预设的迭代次数, 如果未达到, 则根据公式(2)继续更新隶属度矩阵, 如果已达到, 则停止计算, 输出聚类结果和聚类中心, 生成聚类图像。

#### 4. 实验结果及分析

本文基于模糊 C 均值聚类算法对高空天气图像进行了图像分割, 实验所用的高空天气图像来自再分析资料解析出的图像数据集。并通过对照算法在不同噪声情况下的聚类结果, 分析算法对噪声的敏感度。

图 9 中左边的图像是一张原始的高空天气图像, 图 9 右边是基于模糊 C 均值聚类算法对左边图像的分割结果。图 10 和图 11 的左部分图像分别是原始高空天气图像在添加高斯噪声和椒盐噪声之后的图像, 右边分别是模糊 C 均值聚类算法对左边图像进行图像分割之后的结果。从图 9 可以看出模糊 C 均值聚类算法可以将原始高空天气图像划分成为较为清晰的边界, 高空天气图像的各个区域也清晰的分割开。图 10 和图 11 中, 每幅图左边图像加入了高斯噪声和椒盐噪声, 虽然每幅图右边图像的分割结果有变化, 但是天气图的各个区域基本能够显示出来, 但是右下方的区域分割效果相对差些。说明相对于原始图像, 模糊 C 均值聚类算法的分割精度在噪声图像的分割效果有所下降。

为了更为可观的比较算法在不同噪声情况下的分割结果, 使用 DICE 指标来评估聚类结果好坏, 通过计算分割图像与原图像相交面积与原图像面积的比值, 来衡量图像分割效果。DICE 指数取 0 到 1 之间的值, DICE 取值越接近 0, 表示结果越差; 越接近 1, 表示效果越好。除了评估 DICE 指数外, 还对 VOE (体素重叠误差 Volumetric Overlap Error)、RVD (体素相对误差 Relative Volume Difference)、Recall (召回率)指数进行评估。VOE 通过评估原始图像与分割之后的图像两区域的并集和交集的差值与并集之间的比值, 比值越接近于 0, 分割效果越好。RVD 表示分割之后的图像与原始图像在面积之间的差异, 指数范围同 VOE, 越接近 0 表示结果越好。Recall (召回率), 通过计算分割出的像素点个数占于原始图像中像素个数的比值, 指数范围同 DICE, 越接近 1 表示结果越好。如表 1 所示。表 1 展示出了对应于图 9、图 10、图 11 中的每个测试图像分割结果的多个评价指数, 从表中可以看出, 模糊 C 均值聚类对高空天气图像具有很好的分割效果。

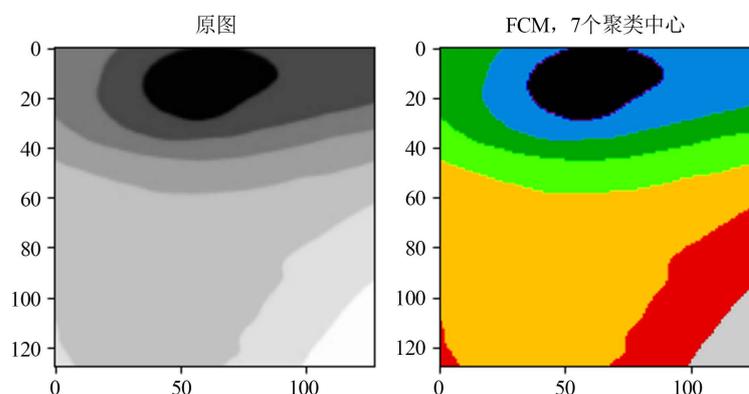


Figure 9. Original image and clustering segmentation image

图 9. 原始图像与聚类分割图像

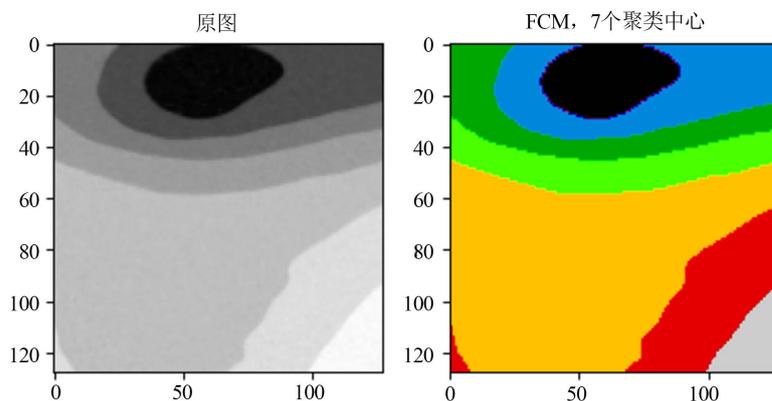


Figure 10. Gaussian noise image and clustering segmentation image  
图 10. 高斯噪声图像与聚类分割图像

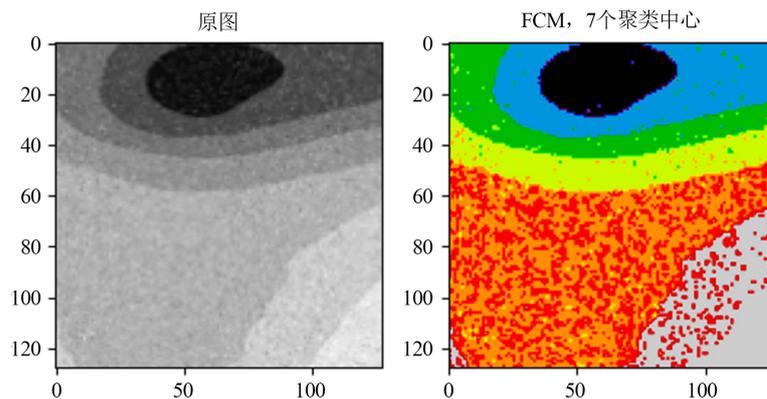


Figure 11. Salt and pepper noise image and clustering segmentation image  
图 11. 椒盐噪声图像与聚类分割图像

Table 1. The value of FCM image segmentation under different evaluation indexes  
表 1. FCM 图像分割在不同评价指数下的值

图像	评价指数			
	DICE	VOE	RVD	Recall
原图像	0.99900	-0.00098	-0.00098	0.99850
高斯噪声图像	0.99360	-0.01015	-0.01009	0.99870
椒盐噪声图像	0.99510	-0.00684	-0.00681	0.99850

### 5. 结论

本文主要对模糊 C 均值聚算法的原理、实现步骤和图像分割流程进行了说明, 将历史再分析资料解析为图像数据, 并结合模糊 C 均值聚类算法对解析出的天气图像以及添加了高斯噪声和椒盐噪声的天气图像进行了图像分割。实验结果表明: 模糊 C 均值聚类算法能够将天气图像等值线区域清晰地分割出来; 对于添加了高斯噪声和椒盐噪声的天气图像, 分割效果受到的影响较小, 说明即使有噪声的干扰, 模糊 C 均值聚类算法的在图像分割中的表现效果也比较稳定。未来的研究工作将继续调整扩大实验研究范围, 得到更为广泛的区域分布数据, 通过自适应区域分布函数来实现不同图像数据下的动态分割。

## 参考文献

- [1] 周莉莉, 姜枫. 图像分割方法综述研究[J]. 计算机应用研究, 2017, 34(7): 1921-1928.
- [2] 赵祥宇, 陈沫涵. 基于聚类的图像分割方法综述[J]. 信息技术, 2018(6): 92-94+99.
- [3] 陈建文, 徐冠雷. 基于 K-Means 的时间季节反演识别方法[J]. 图像与信号处理, 2018, 7(1): 57-64.
- [4] Ngo, L.T., Mai, D.S. and Pedrycz, W. (2015) Semi-Supervising Interval Type-2 Fuzzy C-Means Clustering with Spatial Information for Multi-Spectral Satellite Image Classification and Change Detection. *Computers & Geosciences*, **83**, 1-16. <https://doi.org/10.1016/j.cageo.2015.06.011>
- [5] 胡学刚, 段瑶. 基于 FCM 聚类的自适应彩色图像分割算法[J]. 计算机工程与设计, 2018, 39(7): 1984-1989.
- [6] 胡学刚, 严思奇. 基于 FCM 聚类的图像分割算法[J]. 计算机工程与设计, 2018, 39(1): 159-164.
- [7] 冯仁光. 基于自适应权重 FCM 的浮选泡沫图像分割及应用[D]: [硕士学位论文]. 大连: 大连理工大学, 2014.
- [8] Feng, L., Li, H., Gao, Y., *et al.* (2020) A Color Image Segmentation Method Based on Region Salient Color and Fuzzy C-Means Algorithm. *Circuits, Systems, and Signal Processing*, **39**, 586-610. <https://doi.org/10.1007/s00034-019-01126-w>
- [9] Wu, S., Pang, Y., Shao, S., *et al.* (2018) Advanced Fuzzy C-Means Algorithm Based on Local Density and Distance. *Journal of Shanghai Jiaotong University (Science)*, **23**, 636-642. <https://doi.org/10.1007/s12204-018-1993-y>
- [10] 沉晓. 基于模糊 C 均值聚类的脑部 MR 图像分割算法研究[D]: [硕士学位论文]. 南京: 南京邮电大学, 2019.
- [11] Huang, H., Meng, F., Zhou, S., *et al.* (2019) Brain Image Segmentation Based on FCM Clustering Algorithm and Rough Set. *IEEE Access*, **7**, 12386-12396. <https://doi.org/10.1109/ACCESS.2019.2893063>
- [12] 雷涛, 张肖, 加小红, 刘侍刚, 张艳宁. 基于模糊聚类的图像分割研究进展[J]. 电子学报, 2019, 47(8): 1776-1791.
- [13] 许树成. 基于模糊聚类的图像分割算法的研究[D]: [硕士学位论文]. 北京: 北京邮电大学, 2018.
- [14] 胡加亮. 基于模糊聚类算法的图像分割研究[D]: [硕士学位论文]. 上海: 华东师范大学, 2019.
- [15] 刘野, 黄贤英, 刘文星, 朱小飞, 李昭平. 基于自适应噪声添加的防御对抗样本算法[J]. 计算机应用研究, 2021, 38(3): 764-769.