دیتاست‌های مورد استفاده در خوشه‌بندی

|  |
| --- |
| تهیه کننده گزارش: سید مصطفی غضنفری |
| شماره گزارش: |
| تاریخ تهیه گزارش: 15/08/1403 |
| تعداد صفحات: صفحه |
| بر اساس داده‌های : |
| نام فایل کد مربوطه : |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| جدول Revision History | | | | |
| Rev | **شرح تغییرات** | **تهیه کننده گزارش** | **تأییدکنندگان** | **تاریخ** |
| 1 |  |  |  |  |
| 2 |  |  |  |  |
| 3 |  |  |  |  |
| 4 |  |  |  |  |
| 5 |  |  |  |  |
| 6 |  |  |  |  |
| 7 |  |  |  |  |
| 8 |  |  |  |  |
| 9 |  |  |  |  |
| 10 |  |  |  |  |

Table of Contents

[1 دیتاست‌ها 3](#_Toc181707984)

[1.1 CIFAR-10 3](#_Toc181707985)

[1.2 MNIST 4](#_Toc181707986)

[1.3 Fashion-MNIST 5](#_Toc181707987)

[1.4 Tiny ImageNet 6](#_Toc181707988)

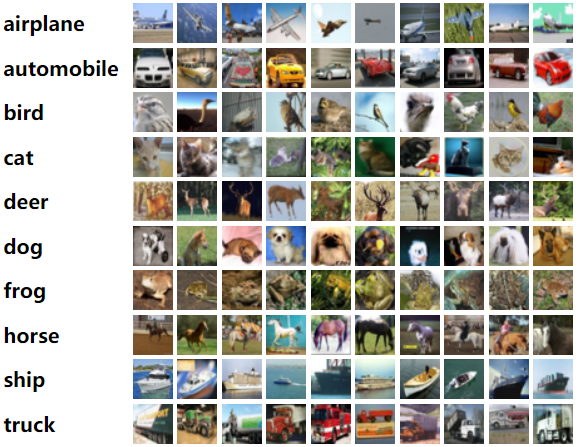
[1.5 STL-10 7](#_Toc181707989)

[2 نکات 8](#_Toc181707990)

# دیتاست‌ها

## CIFAR-10

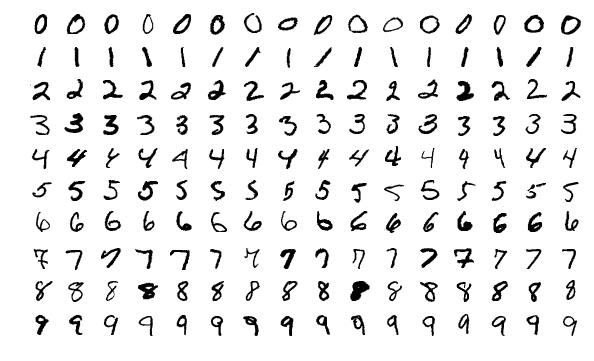
این دیتاست از 50000 تصویر برای آموزش و 10000 تصویر برای تست تشکیل شده است. ابعاد تصاویر در این دیتاست 32\*32 است. این دیتاست از 10 کلاس تشکیل شده است.



در <https://paperswithcode.com/sota/image-clustering-on-cifar-10> بالاترین نتایجی که بر روی این دیتاست در حوزه image clustering به دست آمده قابل مشاهده است.

## MNIST

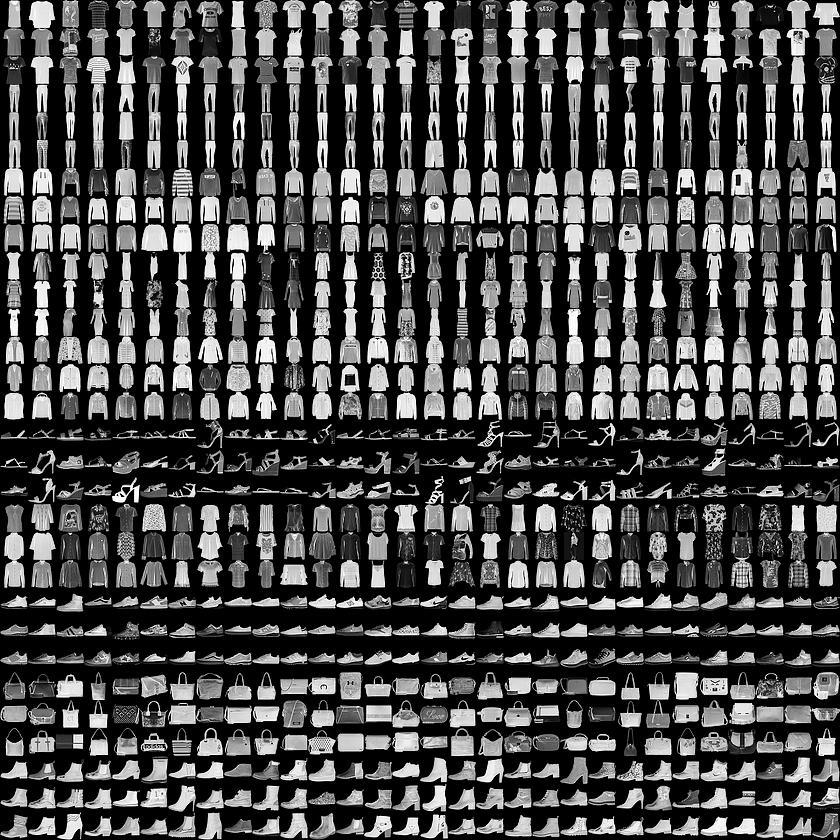
این دیتاست از 60000 تصویر برای آموزش و 10000 تصویر برای تست تشکیل شده است. ابعاد تصاویر در این دیتاست 28\*28 است. این دیتاست از 10 کلاس تشکیل شده است.



در <https://paperswithcode.com/sota/image-clustering-on-mnist-full> بالاترین نتایجی که بر روی این دیتاست در حوزه image clustering به دست آمده قابل مشاهده است.

## Fashion-MNIST

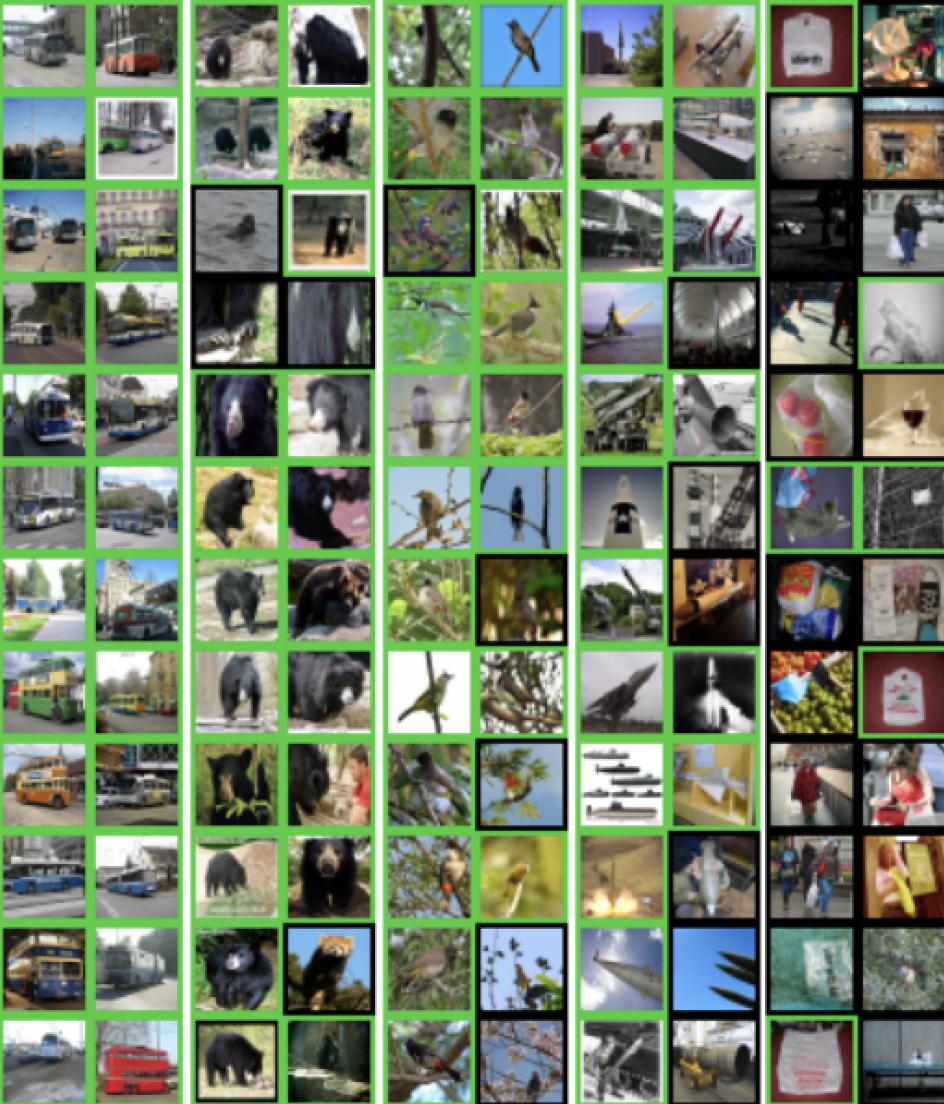
این دیتاست از 60000 تصویر برای آموزش و 10000 تصویر برای تست تشکیل شده است. ابعاد تصاویر در این دیتاست 28\*28 است. این دیتاست از 10 کلاس تشکیل شده است.



در <https://paperswithcode.com/sota/image-clustering-on-fashion-mnist> بالاترین نتایجی که بر روی این دیتاست در حوزه image clustering به دست آمده قابل مشاهده است.

## Tiny ImageNet

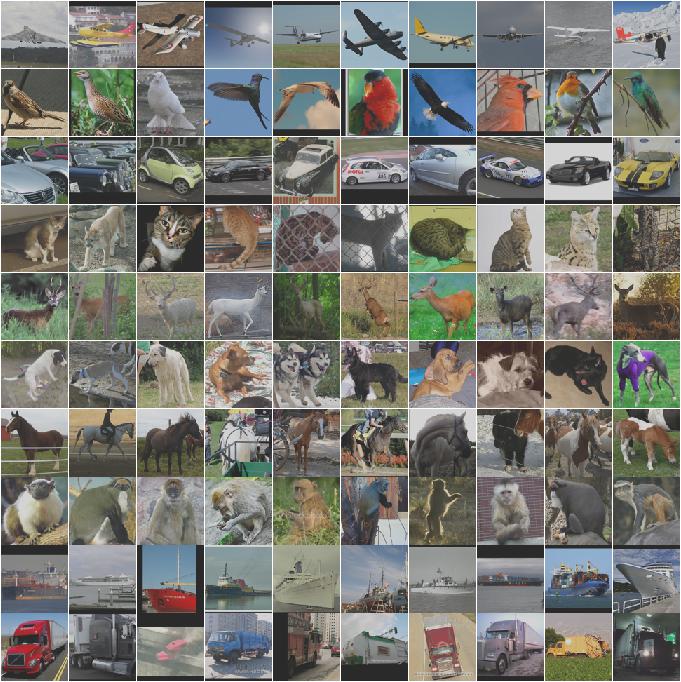
این دیتاست از 100000 تصویر برای آموزش و 10000 تصویر برای ارزیابی و 10000 تصویر برای تست تشکیل شده است. ابعاد تصاویر در این دیتاست 64\*64 است. این دیتاست از 200 کلاس تشکیل شده است.



در <https://paperswithcode.com/sota/image-clustering-on-tiny-imagenet> بالاترین نتایجی که بر روی این دیتاست در حوزه image clustering به دست آمده قابل مشاهده است.

## STL-10

این دیتاست از 100000 تصویر بدون لیبل و همچنین 5000 تصویر برای آموزش و 8000 تصویر برای تست تشکیل شده است. ابعاد تصاویر در این دیتاست 96\*96 است. این دیتاست از 10 کلاس تشکیل شده است.



در <https://paperswithcode.com/sota/image-clustering-on-stl-10> بالاترین نتایجی که بر روی این دیتاست در حوزه image clustering به دست آمده قابل مشاهده است.

# نکات

1. برای انجام تسک خوشه بندی با شبکه های عصبی به تعداد داده زیادی نیاز داریم و یکی از عواملی که این روش ها به نتایج خوبی دست پیدا کرده اند این تعداد داده زیاد است.
2. ابعاد تصاویری که در این دیتاست های معروف میباشد حداقل 28\*28 میباشد. در تصاویر با ابعاد کم استخراج ویژگی به درستی انجام نمیگیرد و ممکن است اصلا ویژگی های معنا دار و زیادی وجود نداشته باشد.
3. دقت خوشه‌بندی تقریبا در همه موارد کمتر یا مساوی دسته بندی میباشد.