数据集构建与使用说明

一、训练集（8 种字体，字号选择 2 种，小四和三号，每个字符 16 个样本）

1、宋体：

哈尔滨工业大学的校训是规格严格功夫到家

2、仿宋：

哈尔滨工业大学的校训是规格严格功夫到家

3、微软雅黑：

哈尔滨工业大学的校训是规格严格功夫到家

4、华光隶书：

哈尔滨工业大学的校训是规格严格功夫到家

5、华光姚体：

哈尔滨工业大学的校训是规格严格功夫到家

6、华光彩云：

哈尔滨工业大学的校训是规格严格功夫到家

7、华光行书：

哈尔滨工业大学的校训是规格严格功夫到家

8、华光行楷：

哈尔滨工业大学的校训是规格严格功夫到家

二、测试集（4 种字体，字号选择 2 种，小四和三号，每个字符 8 个样本）

使用了与训练集不同的另外 4 种字体，具体内容略。

三、数据集构建流程（已经提供，无需实现）：

1、基于训练集和测试集的内容要求，打印输出训练集和测试集文档样张，然后基于扫描仪或手机等方式转换为图片格式。

2、自动或手动切割图片中的文字，得到单个字符图像样本。

四、数据集使用

1、将 train_small 和 train_large 进行合并，构成完整训练数据集。根据样本顺序特点，构建每个字符图片的对应标签。

2、提供的字符样本图片是原始灰度图进行了二值化处理后的版本，质量稍有降低。

3、可选步骤：训练数据集可根据需要进行扩充，如增加不同的字号，通过扫描仪、手机等不同方式及不同的采集参数将电子文档或纸质文档转换为图片格式。如果采用了数据扩充，提交文档中必须对具体方式进行说明。

4、可选步骤：训练数据集也可进行一定的数据增强，如图像变换或增加噪声等。如果采用了数据增强，提交文档中必须对具体方式进行说明。

5、训练集样本进行扩充或增强时，不得使用测试集中的字体类型或测试集中的样本。

6、测试集样本不得做任何改动。

文字样本样张制作

训练集 1

Train_small

文字版

哈尔滨工业大学的校训是规格严格功夫到家

哈尔滨工业大学的校训是规格严格功夫到家

哈尔滨工业大学的校训是规格严格功夫到家

哈尔滨工业大学的校训是规格严格功夫到家

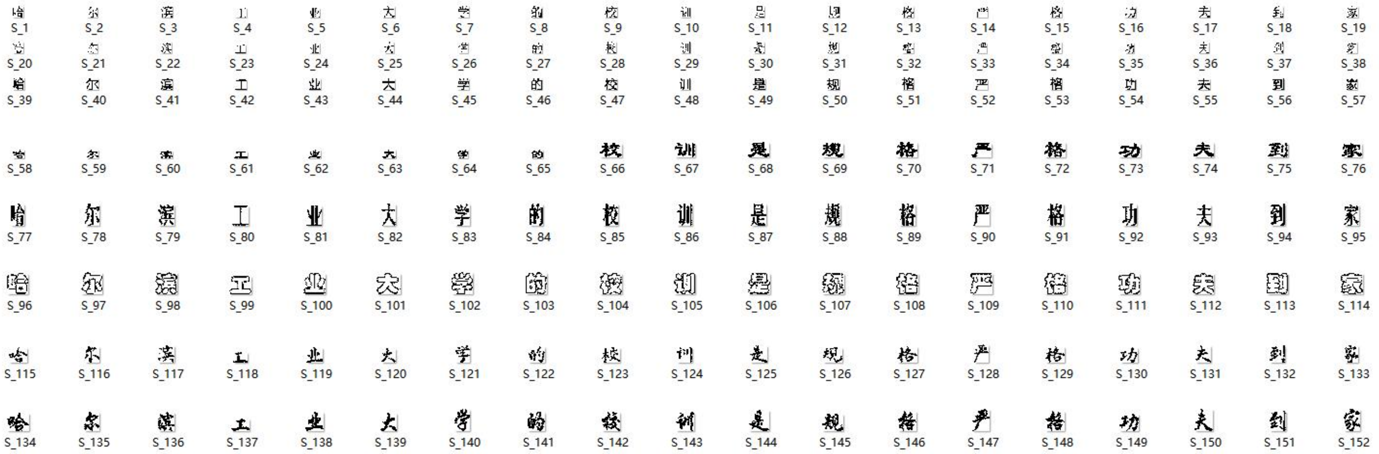
哈尔滨工业大学的校训是规格严格功夫到家

哈尔滨工业大学的校训是规格严格功夫到家

哈尔滨工业大学的校训是规格严格功夫到家

哈尔滨工业大学的校训是规格严格功夫到家

图片版：（二值8位灰度图）



训练集 2

Train_Large

文字版

哈 尔 滨 工 业 大 学 的 校 训 是 规 格 严 格 功 夫 到 家

哈 尔 滨 工 业 大 学 的 校 训 是 规 格 严 格 功 夫 到 家

哈 尔 滨 工 业 大 学 的 校 训 是 规 格 严 格 功 夫 到 家

哈 尔 滨 工 业 大 学 的 校 训 是 规 格 严 格 功 夫 到 家

哈 尔 滨 工 业 大 学 的 校 训 是 规 格 严 格 功 夫 到 家

哈 尔 滨 工 业 大 学 的 校 训 是 规 格 严 格 功 夫 到 家

哈 尔 滨 工 业 大 学 的 校 训 是 规 格 严 格 功 夫 到 家

哈 尔 滨 工 业 大 学 的 校 训 是 规 格 严 格 功 夫 到 家

图片版（二值8位灰度图）

哈 L_1	尔 L_2	滨 L_3	工 L_4	业 L_5	大 L_6	学 L_7	的 L_8	校 L_9	训 L_10	是 L_11	规 L_12	格 L_13	严 L_14	格 L_15	功 L_16	夫 L_17	到 L_18	家 L_19
哈 L_20	尔 L_21	滨 L_22	工 L_23	业 L_24	大 L_25	学 L_26	的 L_27	校 L_28	训 L_29	是 L_30	规 L_31	格 L_32	严 L_33	格 L_34	功 L_35	夫 L_36	到 L_37	家 L_38
哈 L_39	尔 L_40	滨 L_41	工 L_42	业 L_43	大 L_44	学 L_45	的 L_46	校 L_47	训 L_48	是 L_49	规 L_50	格 L_51	严 L_52	格 L_53	功 L_54	夫 L_55	到 L_56	家 L_57
哈 L_58	尔 L_59	滨 L_60	工 L_61	业 L_62	大 L_63	学 L_64	的 L_65	校 L_66	训 L_67	是 L_68	规 L_69	格 L_70	严 L_71	格 L_72	功 L_73	夫 L_74	到 L_75	家 L_76
哈 L_77	尔 L_78	滨 L_79	工 L_80	业 L_81	大 L_82	学 L_83	的 L_84	校 L_85	训 L_86	是 L_87	规 L_88	格 L_89	严 L_90	格 L_91	功 L_92	夫 L_93	到 L_94	家 L_95
哈 L_96	尔 L_97	滨 L_98	工 L_99	业 L_100	大 L_101	学 L_102	的 L_103	校 L_104	训 L_105	是 L_106	规 L_107	格 L_108	严 L_109	格 L_110	功 L_111	夫 L_112	到 L_113	家 L_114
哈 L_115	尔 L_116	滨 L_117	工 L_118	业 L_119	大 L_120	学 L_121	的 L_122	校 L_123	训 L_124	是 L_125	规 L_126	格 L_127	严 L_128	格 L_129	功 L_130	夫 L_131	到 L_132	家 L_133
哈 L_134	尔 L_135	滨 L_136	工 L_137	业 L_138	大 L_139	学 L_140	的 L_141	校 L_142	训 L_143	是 L_144	规 L_145	格 L_146	严 L_147	格 L_148	功 L_149	夫 L_150	到 L_151	家 L_152

测试集

文字版

略

图片版（二值 8 位灰度图）

哈	尔	滨	工	业	大	学	的	校	训	是	规	格	严	格	功	夫	到	家
T_1	T_2	T_3	T_4	T_5	T_6	T_7	T_8	T_9	T_10	T_11	T_12	T_13	T_14	T_15	T_16	T_17	T_18	T_19
哈	尔	滨	工	业	大	学	的	校	训	是	规	格	严	格	功	夫	到	家
T_20	T_21	T_22	T_23	T_24	T_25	T_26	T_27	T_28	T_29	T_30	T_31	T_32	T_33	T_34	T_35	T_36	T_37	T_38
哈	尔	滨	工	业	大	学	的	校	训	是	规	格	严	格	功	夫	到	家
T_39	T_40	T_41	T_42	T_43	T_44	T_45	T_46	T_47	T_48	T_49	T_50	T_51	T_52	T_53	T_54	T_55	T_56	T_57
哈	尔	滨	工	业	大	学	的	校	训	是	规	格	严	格	功	夫	到	家
T_58	T_59	T_60	T_61	T_62	T_63	T_64	T_65	T_66	T_67	T_68	T_69	T_70	T_71	T_72	T_73	T_74	T_75	T_76
哈	尔	滨	工	业	大	学	的	校	训	是	规	格	严	格	功	夫	到	家
T_77	T_78	T_79	T_80	T_81	T_82	T_83	T_84	T_85	T_86	T_87	T_88	T_89	T_90	T_91	T_92	T_93	T_94	T_95
哈	尔	滨	工	业	大	学	的	校	训	是	规	格	严	格	功	夫	到	家
T_96	T_97	T_98	T_99	T_100	T_101	T_102	T_103	T_104	T_105	T_106	T_107	T_108	T_109	T_110	T_111	T_112	T_113	T_114
哈	尔	滨	工	业	大	学	的	校	训	是	规	格	严	格	功	夫	到	家
T_115	T_116	T_117	T_118	T_119	T_120	T_121	T_122	T_123	T_124	T_125	T_126	T_127	T_128	T_129	T_130	T_131	T_132	T_133
哈	尔	滨	工	业	大	学	的	校	训	是	规	格	严	格	功	夫	到	家
T_134	T_135	T_136	T_137	T_138	T_139	T_140	T_141	T_142	T_143	T_144	T_145	T_146	T_147	T_148	T_149	T_150	T_151	T_152