

A SUPPLEMENTARY MATERIALS

A.1 ACCESS INSTRUCTIONS

Ins-MMBench and the code for the evaluation experiments will be available.

A.2 DATA FORMAT

The INS-MMBench dataset is formatted as a multiple-choice visual question and is stored in a .tsv file. Each entry includes the following fields: *index*, *question*, *answer*, *A*, *B*, *C*, *D*, *image*, *category*, *l2 category*, and *l3 category*. These fields represent the instance index, the question associated with the image, the correct answer, the four answer options (A, B, C, D), the base64-encoded image, the specific type of insurance task, the meta-task name, and the basic task name, respectively.

A.3 HOSTING PLAN

INS-MMBench will be hosted on Github for the foreseeable future. Should that change, we will create an issue on the GitHub repository.

B DATASHEET

B.1 MOTIVATION

For what purpose was the dataset created?

The purpose of creating INS-MMBench was to comprehensively evaluate the capabilities of LVLMS in the insurance field. It includes 12 meta-tasks and 22 fundamental tasks. It is the first comprehensive multimodal benchmark in the insurance field.

B.2 COMPOSITION

What do the instances that comprise the dataset represent (e.g., documents, photos, people, countries)?

The instance of INS-MMBench represents a multiple-choice visual question, including the instance index, the question corresponding to the image, the correct answer, four options A, B, C, and D, the base64 encoding of the image, the insurance type to which the task belongs, the meta-task name, and the basic task name.

Does the dataset contain all possible instances or is it a sample (not necessarily random) of instances from a larger set?

Given the expense associated with LVLMS and the need for convenient model evaluation, the image data for each of our core tasks was randomly selected from a larger dataset.

What data does each instance consist of?

Each instance in the INS-MMBench dataset includes the following components: *index*, *question*, *answer*, *A*, *B*, *C*, *D*, *image*, *category*, *l2 category*, and *l3 category*. These components correspond respectively to the instance index, the question related to the image, the correct answer, four answer options (A, B, C, and D), the base64-encoded image, the type of insurance related to the task, the meta-task name, and the basic task name.

Is there a label or target associated with each instance?

Yes.

Is any information missing from individual instances?

No.

Are relationships between individual instances made explicit (e.g., users' movie ratings, social network links)?

N/A.

Are there recommended data splits (e.g., training, development/validation, testing)?

No.

Are there any errors, sources of noise, or redundancies in the dataset?

The authors are not aware of any errors in the dataset.

Does the dataset contain data that might be considered confidential (e.g., data that is protected by legal privilege or by doctor-patient confidentiality, data that includes the content of individuals' non-public communications)?

No.

Does the dataset contain data that, if viewed directly, might be offensive, insulting, threatening, or might otherwise cause anxiety?

No.

Does the dataset relate to people?

No.

B.3 COLLECTION PROCESS

How was the data associated with each instance acquired?

First, we collected datasets from multiple public sources. Next, we performed manual filtering and random sampling of the datasets, followed by the necessary data processing. Finally, both manual effort and GPT-4o were utilized to construct task questions and multiple-choice options, creating a multi-choice visual question dataset.

What mechanisms or procedures were used to collect the data (e.g., hardware apparatus or sensor, manual human curation, software program, software API)?

We used both manual human curation and the GPT-4o API.

If the dataset is a sample from a larger set, what was the sampling strategy (e.g., deterministic, probabilistic with specific sampling probabilities)?

We employed the uniform sampling according to different label distributions.

Who was involved in the data collection process (e.g., students, crowdworkers, contractors) and how were they compensated (e.g., how much were crowdworkers paid)?

We did not employ external crowdworkers or contractors to construct INS-MMBench.

Over what timeframe was the data collected?

We completed data collection in May 2024.

Were any ethical review processes conducted (e.g., by an institutional review board)?

No.

B.4 PREPROCESSING/CLEANING/LABELING

Was any preprocessing/cleaning/labeling of the data done (e.g., discretization or bucketing, tokenization, part-of-speech tagging, SIFT feature extraction, removal of instances, processing of missing values)?

Yes, we constructed questions and options for different images using both manual and GPT-4o assistance, and evenly distributed the answers.

B.5 USES

Has the dataset been used for any tasks already?

Yes, we have evaluated 10 LVLMs on this dataset.

Is there a repository that links to any or all papers or systems that use the dataset?

No.

What (other) tasks could the dataset be used for?

Beyond assessing the capabilities of LVLMs in the insurance sector, this dataset can also serve as fine-tuning training data to enhance LVLM performance.

Is there anything about the composition of the dataset or the way it was collected and preprocessed/cleaned/labeled that might impact future uses?

The task scope and number of instances in this dataset are limited. However, future usage and research could expand and enhance based on this foundation.

Are there tasks for which the dataset should not be used?

No.

B.6 DISTRIBUTION

Will the dataset be distributed to third parties outside of the entity (e.g., company, institution, organization) on behalf of which the dataset was created?

Yes.

How will the dataset will be distributed (e.g., tarball on website, API, GitHub)?

Ins-MMBench will be distributed at the Github repository.

When will the dataset be distributed?

The dataset will be made available upon acceptance.

C CASES

To offer a detailed view of the task settings in INS-MMBench, we have selected sample cases for each core task and present responses from GPT-4o, Gemini 1.5 Flash, and Qwen-VL-Max in this section.

License plate recognition

Prompt:
 What is the license plate number of the vehicle in the image?
 Options:
 A. NEWJOB
 B. NEWLOB
 C. N3WJOB
 D. MEWJOB

Correct answer: A



✓

GPT-4o:
 The license plate number of the vehicle in the image is:
 A. NEWJOB

✓

Gemini 1.5 Flash:
 The answer is A. The license plate number is NEWJOB.

✓

Qwen-VL-Max:
 The license plate number of the vehicle in the image is "NEWJOB".

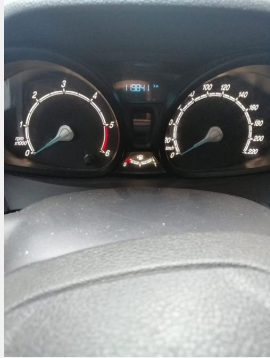
Figure 5: A sample case of license plate recognition and the responses of GPT-4o, GeminiProVision and Qwen-VL-Max.

1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040
1041
1042
1043
1044
1045
1046
1047
1048
1049
1050
1051
1052
1053
1054
1055
1056
1057
1058
1059
1060
1061
1062
1063
1064
1065
1066
1067
1068
1069
1070
1071
1072
1073
1074
1075
1076
1077
1078
1079

Vehicle milage reading

Prompt:
What is the mileage reading of the vehicle in the image?
Options:
A. 119841
B. 118941
C. 119481
D. 191841

Correct answer: A



GPT-4o:
The mileage reading of the vehicle in the image is 119841.
So, the correct answer is:
A. 119841

Gemini 1.5 Flash:
The answer is A. 119841.

Qwen-VL-Max:
The mileage reading of the vehicle in the image is option B, 118941 kilometers.

Figure 6: A sample case of vehicle mileage reading and the responses of GPT-4o, GeminiProVision and Qwen-VL-Max.

Vehicle warning indicator recognition

Prompt:

Which indicator icons are lit on the vehicle dashboard in the image?

Options:

- A. Door open, Low Fuel
- B. Check Engine, High Beam
- C. Tire Pressure, Battery
- D. Engine Oil, ABS



Correct answer: A



GPT-4o:

The correct answer is:
A. Door open, Low Fuel



Gemini 1.5 Flash:

The correct answer is C. Tire Pressure, Battery.
The image shows two indicator icons lit up on the dashboard: a tire pressure icon and a battery icon.



Qwen-VL-Max:

The correct answer is A. Door open, Low Fuel. In the image of the vehicle dashboard, there are two indicator icons that are lit up. The first icon indicates that one or more doors are open, and the second icon indicates that the fuel level is low. These symbols serve as important reminders for the driver to ensure safety and proper vehicle maintenance.

Figure 7: A sample case of vehicle warning indicator recognition and the responses of GPT-4o, GeminiProVision and Qwen-VL-Max.

1134
1135
1136
1137
1138
1139
1140
1141
1142
1143
1144
1145
1146
1147
1148
1149
1150
1151
1152
1153
1154
1155
1156
1157
1158
1159
1160
1161
1162
1163
1164
1165
1166
1167
1168
1169
1170
1171
1172
1173
1174
1175
1176
1177
1178
1179
1180
1181
1182
1183
1184
1185
1186
1187

Vehicle make and model identification

Prompt:

What is the make and model of the vehicle in the image?

Options:

- A. Jaguar XKR-S Convertible 2012
- B. Bentley Continental GT Convertible 2012
- C. Rolls-Royce Phantom Drophead Coupe Convertible 2012U
- D. Aston Martin DB9 Volante Convertible 2012



Correct answer: C



GPT-4o:

The vehicle in the image is a Rolls-Royce Phantom Drophead Coupe Convertible 2012. So, the correct answer is:

C. Rolls-Royce Phantom Drophead Coupe Convertible 2012



Gemini 1.5 Flash:

The answer is C. Rolls-Royce Phantom Drophead Coupe Convertible 2012.



Qwen-VL-Max:

The make and model of the vehicle in the image is C. Rolls-Royce Phantom Drophead Coupe Convertible 2012.

Figure 8: A sample case of vehicle make and model identification and the responses of GPT-4o, GeminiProVision and Qwen-VL-Max.

Prompt:

A. Yes

Correct answer: A



A. Yes



A. Yes.



A. Yes

23

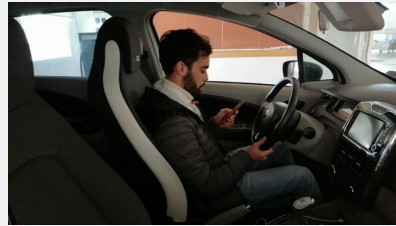
Incar driving behavior detection

Prompt:

What is the behavior of the driver in the image?

Options:

- A. Texting - left
- B. Adjusting mirror - right
- C. Eating - right
- D. Talking on the phone - left



Correct answer: A



GPT-4o:

A. Texting - left



Gemini 1.5 Flash:

The answer is A. Texting - left. The driver is holding a phone in his left hand and looking at it.



Qwen-VL-Max:

The correct answer is D. Talking on the phone - left. In the image, the driver is sitting in the car and holding his phone to his ear, indicating that he is talking on the phone. He is not texting or adjusting the mirror as suggested by options A and B respectively. Option C, eating, does not apply since there is no visible food or action related to eating in the image.

Figure 10: A sample case of incar driving behavior detection and the responses of GPT-4o, GeminiProVision and Qwen-VL-Max.

1296
1297
1298
1299
1300
1301
1302
1303
1304
1305
1306
1307
1308
1309
1310
1311
1312
1313
1314
1315
1316
1317
1318
1319
1320
1321
1322
1323
1324
1325
1326
1327
1328
1329
1330
1331
1332
1333
1334
1335
1336
1337
1338
1339
1340
1341
1342
1343
1344
1345
1346
1347
1348
1349

Vehicle damage part detection

Prompt:

What are the damaged parts of the vehicle in the image?

Options:

- A. Bonnet-dent
- B. Door-scratch
- C. Bumper-dent
- D. Bonnet-crack



Correct answer: A



GPT-4o:

C. Bumper-dent



Gemini 1.5 Flash:

The answer is A. Bonnet-dent.

The image shows a dent on the bonnet of the vehicle.



Qwen-VL-Max:

The correct answer is A. Bonnet-dent. In the image, there is a blue BMW car with visible damage on its front end. The hood (bonnet) has a dent and some scratches, while one of the grills appears to be missing or broken. The bumper also seems to have some dents and scratches. The license plate holder is bent, and the license plate itself is slightly crooked.

Figure 11: A sample case of vehicle damage part detection and the responses of GPT-4o, GeminiPro-Vision and Qwen-VL-Max.

1350
1351
1352
1353
1354
1355
1356
1357
1358
1359
1360
1361
1362
1363
1364
1365
1366
1367
1368
1369
1370
1371
1372
1373
1374
1375
1376
1377
1378
1379
1380
1381
1382
1383
1384
1385
1386
1387
1388
1389
1390
1391
1392
1393
1394
1395
1396
1397
1398
1399
1400
1401
1402
1403

Vehicle damage type detection

Prompt:

What are the damaged types of the vehicle in the image?

Options:

- A. Broken Headlight
- B. Dented Door
- C. Scratched Paint
- D. Tire Flat



Correct answer: D



GPT-4o:

The correct answer is:
D. Tire Flat



Gemini 1.5 Flash:

The answer is D. Tire Flat. The tire is clearly deflated and the car is resting on the rim.



Qwen-VL-Max:

The correct answer is D. Tire Flat. In the image, there is a white car with a flat tire on the front left side. The tire appears to be completely deflated and rests against the curb. There is no visible damage to the headlight or door, nor any noticeable scratches on the paint. Therefore, the only visible damage to the vehicle in this image is the flat tire.

Figure 12: A sample case of vehicle damage type detection and the responses of GPT-4o, GeminiPro-Vision and Qwen-VL-Max.

1404
1405
1406
1407
1408
1409
1410
1411
1412
1413
1414
1415
1416
1417
1418
1419
1420
1421
1422
1423
1424
1425
1426
1427
1428
1429
1430
1431
1432
1433
1434
1435
1436
1437
1438
1439
1440
1441
1442
1443
1444
1445
1446
1447
1448
1449
1450
1451
1452
1453
1454
1455
1456
1457

Vehicle damage severity detection

Prompt:
What is the damaged severity level of the vehicle in the image?
Options:
A. severe
B. moderate
C. no_accident
D. minor



Correct answer: D



GPT-4o:
The correct answer is:
D. minor
The vehicle in the image has some visible damage, particularly around the front left fender and headlight area, but it does not appear to be severe.



Gemini 1.5 Flash:
The answer is B. moderate.
The vehicle has visible damage to the front bumper and fender, but the overall structure of the vehicle appears intact. This suggests moderate damage.



Qwen-VL-Max:
The correct answer is D. minor. Based on the image, there appears to be some damage to the front bumper and possibly the headlights of the white car. However, the overall damage seems to be relatively minor compared to more severe accidents or collisions. Therefore, option D, minor, would be the most appropriate description for the damage severity level of the vehicle in the image.

Figure 13: A sample case of vehicle damage severity detection and the responses of GPT-4o, GeminiProVision and Qwen-VL-Max.

1458
1459
1460
1461
1462
1463
1464
1465
1466
1467
1468
1469
1470
1471
1472
1473
1474
1475
1476
1477
1478
1479
1480
1481
1482
1483
1484
1485
1486
1487
1488
1489
1490
1491
1492
1493
1494
1495
1496
1497
1498
1499
1500
1501
1502
1503
1504
1505
1506
1507
1508
1509
1510
1511

Roof condition assessment

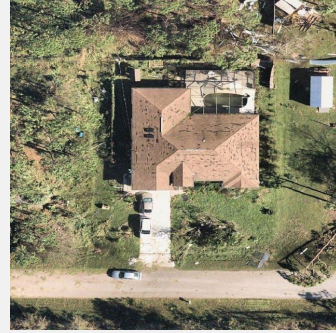
Prompt:

What is the condition of the roof of the house in the image?

Options:

A. Damage

B. Not damage



Correct answer: A



GPT-4o:

A. Damage



Gemini 1.5 Flash:

A. Damage



Qwen-VL-Max:

A. Damage

Figure 14: A sample case of roof condition assessment and the responses of GPT-4o, GeminiProVision and Qwen-VL-Max.

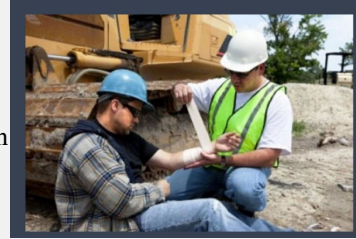
Workplace risk assessment

Prompt:

What's the risk condition of the workplace in the image?

Options:

- A. There is risk of not wearing helmet in the image
- B. There is risk of not wearing helmet and vest in the image
- C. There is risk of not wearing vest in the image
- D. There is no risk in the image



Correct answer: C



GPT-4o:

- A. There is risk of not wearing helmet in the image



Gemini 1.5 Flash:

There is risk of not wearing vest in the image



Qwen-VL-Max:

C. There is risk of not wearing vest in the image.
In the image, one worker is seen wearing a hard hat and helping another worker who appears to be injured. However, neither of them is wearing a safety vest, which poses a risk on the construction site. Therefore, option C is the correct answer.

Figure 15: A sample case of workplace risk assessment and the responses of GPT-4o, GeminiProVision and Qwen-VL-Max.

1566
1567
1568
1569
1570
1571
1572
1573
1574
1575
1576
1577
1578
1579
1580
1581
1582
1583
1584
1585
1586
1587
1588
1589
1590
1591
1592
1593
1594
1595
1596
1597
1598
1599
1600
1601
1602
1603
1604
1605
1606
1607
1608
1609
1610
1611
1612
1613
1614
1615
1616
1617
1618
1619

House fire detection

Prompt:
Is there fire in the house in the image?
Options:
A. no fire
B. fire



Correct answer: B

GPT-4o:
A. no fire

Gemini 1.5 Flash:
A. no fire

Qwen-VL-Max:
Option A: No fire.
There is no evidence of a fire in the image. The room appears to be well-lit and there are no signs of smoke or flames.

Figure 16: A sample case of hosue fire detection and the responses of GPT-4o, GeminiProVision and Qwen-VL-Max.

1620
1621
1622
1623
1624
1625
1626
1627
1628
1629
1630
1631
1632
1633
1634
1635
1636
1637
1638
1639
1640
1641
1642
1643
1644
1645
1646
1647
1648
1649
1650
1651
1652
1653
1654
1655
1656
1657
1658
1659
1660
1661
1662
1663
1664
1665
1666
1667
1668
1669
1670
1671
1672
1673

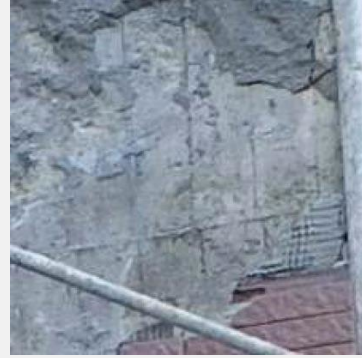
House damage type detection

Prompt:

What type of damage is done to the house in the image?

Options:

- A. Combined Damage
- B. Flexural
- C. Shear



Correct answer: A



GPT-4o:

The damage shown in the image appears to be characterized by diagonal cracks and a pattern that suggests a sliding or shearing motion. This type of damage is typically associated with shear forces. The correct answer is:
C. Shear



Gemini 1.5 Flash:

The answer is A. Combined Damage.
The image shows cracks in the wall, which are a sign of both flexural and shear damage. Flexural damage occurs when a material is bent or deformed, while shear damage occurs when a material is subjected to a force that is parallel to its surface. In this case, the cracks are likely caused by a combination of these two types of damage.



Qwen-VL-Max:

The correct answer is A. Combined Damage.
Based on the image provided, it appears that there is combined damage to the house's exterior wall. The wall shows signs of cracking and peeling, which could be due to a combination of factors such as age, weathering, or structural issues. The presence of a metal mesh suggests that repairs have been attempted, but the damage still remains visible. Therefore, option A, Combined Damage, best describes the type of damage seen in the image.

Figure 17: A sample case of house damage type detection and the responses of GPT-4o, GeminiPro-Vision and Qwen-VL-Max.

1674
1675
1676
1677
1678
1679
1680
1681
1682
1683
1684
1685
1686
1687
1688
1689
1690
1691
1692
1693
1694
1695
1696
1697
1698
1699
1700
1701
1702
1703
1704
1705
1706
1707
1708
1709
1710
1711
1712
1713
1714
1715
1716
1717
1718
1719
1720
1721
1722
1723
1724
1725
1726
1727

House damage level detection

Prompt:

What is the level of damage to the house in the image?

Options:

- A. Undamage
- B. Minor Damage
- C. Moderate Damage
- D. Heavy Damage



Correct answer: C



GPT-4o:

The correct answer is:

C. Moderate Damage

The image shows significant cracks and peeling on the exterior wall, indicating structural issues that are more than minor but not extensive enough to be classified as heavy damage.



Gemini 1.5 Flash:

The answer is C. Moderate Damage.

The image shows a large crack running across the wall of the house, indicating a significant structural issue. However, the house is still standing and the crack doesn't appear to have caused a complete collapse. Therefore, it's not heavy damage, but it's more than just minor damage.



Qwen-VL-Max:

The correct answer based on the image provided is:\n\nC. Moderate Damage\n\nThe image shows a pink building with visible cracks and peeling paint, indicating some degree of structural damage or wear over time. The extent of the damage appears to be moderate rather than minor or heavy.

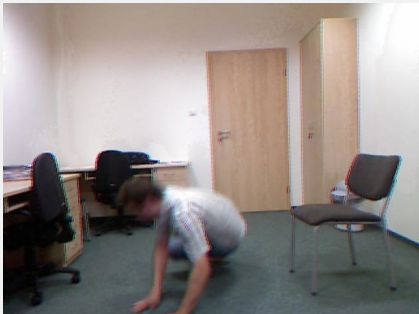
Figure 18: A sample case of house damage level detection and the responses of GPT-4o, GeminiPro-Vision and Qwen-VL-Max.

1728
1729
1730
1731
1732
1733
1734
1735
1736
1737
1738
1739
1740
1741
1742
1743
1744
1745
1746
1747
1748
1749
1750
1751
1752
1753
1754
1755
1756
1757
1758
1759
1760
1761
1762
1763
1764
1765
1766
1767
1768
1769
1770
1771
1772
1773
1774
1775
1776
1777
1778
1779
1780
1781

Fall detection

Prompt:
Does the person in the picture fall down?
Options:
A. fallen
B. not fallen

Correct answer: A



✓ **GPT-4o:**
A. fallen

✓ **Gemini 1.5 Flash:**
A. fallen

✗ **Qwen-VL-Max:**
B. not fallen

Figure 19: A sample case of fall detection and the responses of GPT-4o, GeminiProVision and Qwen-VL-Max.

1782
1783
1784
1785
1786
1787
1788
1789
1790
1791
1792
1793
1794
1795
1796
1797
1798
1799
1800
1801
1802
1803
1804
1805
1806
1807
1808
1809
1810
1811
1812
1813
1814
1815
1816
1817
1818
1819
1820
1821
1822
1823
1824
1825
1826
1827
1828
1829
1830
1831
1832
1833
1834
1835

Health device reading

Prompt:

What is the reading of the blood pressure monitor in the image?

Options:

- A. SYS:130,DIA:70,PULSE:60
- B. SYS:110,DIA:65,PULSE:55
- C. SYS:120,DIA:75,PULSE:50
- D. SYS:120,DIA:67,PULSE:55



Correct answer: D



GPT-4o:

The correct answer is:
D. SYS:120,DIA:67,PULSE:55



Gemini 1.5 Flash:

D. SYS:120,DIA:67,PULSE:55



Qwen-VL-Max:

The correct answer is D. SYS:120,DIA:67,PULSE:55.

Figure 20: A sample case of health device reading and the responses of GPT-4o, GeminiProVision and Qwen-VL-Max.

1836
1837
1838
1839
1840
1841
1842
1843
1844
1845
1846
1847
1848
1849
1850
1851
1852
1853
1854
1855
1856
1857
1858
1859
1860
1861
1862
1863
1864
1865
1866
1867
1868
1869
1870
1871
1872
1873
1874
1875
1876
1877
1878
1879
1880
1881
1882
1883
1884
1885
1886
1887
1888
1889

Medical image abnormality recognition

Prompt:

what is the primary abnormality in this image?

Options:

A. Osteomyelitis

B. Osteogenesis Imperfecta

C. Osteoarthritis

D. Osteosarcoma

Correct answer: B



✓

GPT-4o:

The primary abnormality in the image is consistent with Osteogenesis Imperfecta (Option B). This condition is characterized by bones that break easily, often with little or no apparent cause. The X-ray shows multiple fractures and bone deformities, which are typical features of Osteogenesis Imperfecta.

✗

Gemini 1.5 Flash:

The answer is D. Osteosarcoma. Osteosarcoma is a type of bone cancer that typically affects children and young adults. It is characterized by the formation of malignant tumors in the bone. The image shows a large, destructive tumor in the humerus, which is a classic finding of osteosarcoma.

✓

Qwen-VL-Max:

B. Osteogenesis Imperfecta

Figure 21: A sample case of medical image abnormality detection and the responses of GPT-4o, GeminiProVision and Qwen-VL-Max.

35

1944
1945
1946
1947
1948
1949
1950
1951
1952
1953
1954
1955
1956
1957
1958
1959
1960
1961
1962
1963
1964
1965
1966
1967
1968
1969
1970
1971
1972
1973
1974
1975
1976
1977
1978
1979
1980
1981
1982
1983
1984
1985
1986
1987
1988
1989
1990
1991
1992
1993
1994
1995
1996
1997

Field image crop type identification

Prompt:
What type of crops are in the image?
Options:
A. Jute
B. Sisal
C. Flax
D. Hemp



Correct answer: A

✗

GPT-4o:
D. Hemp

✓

Gemini 1.5 Flash:
The answer is A. Jute.
Jute is a natural fiber obtained from the stem of the jute plant. It is a long, soft, and shiny fiber that is used to make a variety of products, including rope, twine, burlap, and canvas. The image shows a field of jute plants, which are characterized by their tall, slender stems and green leaves.

✗

Qwen-VL-Max:
The correct answer is C. Flax. The image shows flax plants growing in a field.

Figure 23: A sample case of field image crop type identification and the responses of GPT-4o, GeminiProVision and Qwen-VL-Max.

37

1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
2023
2024
2025
2026
2027
2028
2029
2030
2031
2032
2033
2034
2035
2036
2037
2038
2039
2040
2041
2042
2043
2044
2045
2046
2047
2048
2049
2050
2051

Satellite image crop type identification

Prompt:
What type of crops are in the image?
Options:
A. Barley
B. Sorghum
C. Wheat
D. Maize



Correct answer: D

✗

GPT-4o:
The image appears to show a close-up of a crop, but the details are not clear enough to definitively identify the type of crop. However, based on the general appearance, it resembles the structure of wheat. So, the correct answer is:
C. Wheat

✗

Gemini 1.5 Flash:
The answer is C. Wheat.
Wheat crops are characterized by their long, slender leaves and tall stalks. The image shows these features, making it likely that the crops are wheat.

✗

Qwen-VL-Max:
Based on the image provided, it is difficult to determine the exact type of crop as the image is blurry and lacks detail. However, based on the given options, the most likely answer would be C. Wheat, as it is a common cereal grain that is often grown in large fields and has a similar appearance to the作物 in the image. However, without more information or a clearer image, it is impossible to确定ly identify the specific type of crop.

Figure 24: A sample case of satellite image crop type identification and the responses of GPT-4o, GeminiProVision and Qwen-VL-Max.

2052
2053
2054
2055
2056
2057
2058
2059
2060
2061
2062
2063
2064
2065
2066
2067
2068
2069
2070
2071
2072
2073
2074
2075
2076
2077
2078
2079
2080
2081
2082
2083
2084
2085
2086
2087
2088
2089
2090
2091
2092
2093
2094
2095
2096
2097
2098
2099
2100
2101
2102
2103
2104
2105

Crop growth stage recognition

Prompt:

What is the growth stage of the crops in the image?

Options:

- A. Booting
- B. Ripening
- C. Tillering
- D. Flowering



Correct answer: A



GPT-4o:

The growth stage of the crops in the image appears to be "C. Tillering." This stage is characterized by the development of multiple shoots from the base of the plant, which is evident in the dense and lush green appearance of the crop.



Gemini 1.5 Flash:

The answer is C. Tillering.
The crops in the image are in the tillering stage. This is the stage where the plant is producing new shoots from the base of the stem. The shoots are called tillers. The tillering stage is important for the plant because it allows it to produce more leaves and stems, which will eventually produce more grain.



Qwen-VL-Max:

Based on the image, it appears that the crops are at the tillering stage, which is when the plants are developing side shoots and branching out to form a dense canopy. Therefore, the correct answer is C. Tillering.

Figure 25: A sample case of crop growth stage recognition and the responses of GPT-4o, GeminiPro-Vision and Qwen-VL-Max.

2106
2107
2108
2109
2110
2111
2112
2113
2114
2115
2116
2117
2118
2119
2120
2121
2122
2123
2124
2125
2126
2127
2128
2129
2130
2131
2132
2133
2134
2135
2136
2137
2138
2139
2140
2141
2142
2143
2144
2145
2146
2147
2148
2149
2150
2151
2152
2153
2154
2155
2156
2157
2158
2159

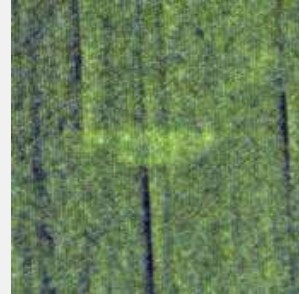
Farmland damage type detection

Prompt:

What type of damage has occurred on the farmland?

Options:

- A. Double plant
- B. Lodging
- C. Erosion
- D. Pest infestation



Correct answer: A



GPT-4o:

B. Lodging



Gemini 1.5 Flash:

The answer is C. Erosion.

Erosion is the process of soil and rock being moved from one place to another by wind, water, or ice. The image shows a pattern of vertical lines, which are likely caused by erosion.



Qwen-VL-Max:

Based on the image provided, it appears that there is lodging damage on the farmland. The plants appear to be leaning or fallen over, which is characteristic of lodging damage caused by strong winds or other factors. Therefore, the correct answer is B. Lodging.

Figure 26: A sample case of farmland damage detection and the responses of GPT-4o, GeminiProVision and Qwen-VL-Max.