

# E DIN EN 18288:2026-06 (E)

Erscheinungsdatum: 2026-05-15

## Artificial Intelligence - Taxonomy of AI tasks in computer vision; English version prEN 18288:2026

---

### Contents

Page

European foreword .....	4
<b>1</b> Scope.....	<b>5</b>
<b>2</b> Normative references .....	<b>5</b>
<b>3</b> Terms and definitions.....	<b>5</b>
<b>3.1</b> Computer vision related terms.....	<b>5</b>
<b>4</b> Abbreviated terms .....	<b>5</b>
<b>5</b> Tasks.....	<b>6</b>
<b>5.1</b> General.....	<b>6</b>
<b>5.2</b> Definition of computer vision task .....	<b>6</b>
<b>5.3</b> Categories of computer vision tasks .....	<b>6</b>
<b>5.4</b> Task variants and versions .....	<b>6</b>
<b>5.5</b> Role of context on computer vision tasks.....	<b>6</b>
<b>6</b> Core computer vision tasks.....	<b>6</b>
<b>6.1</b> General.....	<b>6</b>
<b>6.2</b> Image analysis .....	<b>6</b>
<b>6.2.1</b> General.....	<b>6</b>
<b>6.2.2</b> Object localization.....	<b>7</b>
<b>6.2.3</b> Object classification.....	<b>8</b>
<b>6.2.4</b> Object detection .....	<b>8</b>
<b>6.2.5</b> Semantic segmentation .....	<b>9</b>
<b>6.2.6</b> Instance segmentation.....	<b>9</b>
<b>6.2.7</b> Panoptic segmentation.....	<b>10</b>
<b>6.2.8</b> Image registration.....	<b>10</b>
<b>6.2.9</b> Image similarity .....	<b>11</b>
<b>6.2.10</b> Image classification .....	<b>12</b>
<b>6.2.11</b> Face analysis .....	<b>12</b>
<b>6.2.12</b> Saliency detection .....	<b>12</b>
<b>6.2.13</b> Scene Graph Generation (SGG).....	<b>13</b>
<b>6.2.14</b> Instance identification.....	<b>13</b>
<b>6.2.15</b> Feature extraction.....	<b>14</b>
<b>6.2.16</b> Geometric feature detection.....	<b>14</b>
<b>6.2.17</b> Image stitching.....	<b>15</b>
<b>6.2.18</b> Image retrieval.....	<b>15</b>
<b>6.3</b> Spatial analysis.....	<b>16</b>
<b>6.3.1</b> General.....	<b>16</b>
<b>6.3.2</b> Depth estimation .....	<b>16</b>
<b>6.3.3</b> Object / 3D reconstruction.....	<b>17</b>
<b>6.3.4</b> Object pose estimation .....	<b>18</b>
<b>6.3.5</b> Camera pose estimation.....	<b>19</b>
<b>6.3.6</b> Simultaneous Localization and Mapping (SLAM).....	<b>19</b>
<b>6.4</b> Temporal analysis.....	<b>20</b>
<b>6.4.1</b> General.....	<b>20</b>
<b>6.4.2</b> Multiple Object Tracking (MOT).....	<b>20</b>
<b>6.4.3</b> Action recognition.....	<b>20</b>
<b>6.4.4</b> Event detection .....	<b>21</b>

6.4.5	Object re-identification (Re-ID).....	21
6.4.6	Visual odometry.....	22
6.4.7	Change detection.....	22
6.4.8	Deepfake detection.....	23
6.4.9	Video summarization.....	23
6.4.10	Optical flow estimation.....	23
6.4.11	Motion detection.....	24
6.4.12	Single object tracking.....	24
6.5	Tasks producing image as an output.....	25
6.5.1	General.....	25
6.5.2	Image generation.....	25
6.5.3	Image denoising.....	25
6.5.4	Super-resolution.....	26
6.5.5	Image inpainting.....	26
6.5.6	Image-to-image translation.....	26
6.5.7	Style transfer.....	27
6.5.8	Object swapping.....	27
6.5.9	Novel view synthesis.....	28
6.6	Tasks producing video as an output.....	28
6.6.1	General.....	28
6.6.2	Video enhancement.....	28
6.6.3	Video generation.....	29
6.6.4	Video stabilization.....	29
6.7	Multi-modal tasks.....	30
6.7.1	General.....	30
6.7.2	Visual question answering (VQA).....	30
6.7.3	Image captioning.....	30
6.7.4	Visual grounding.....	31
6.7.5	RGB-depth fusion.....	31
	Bibliography.....	32

## Figures

Figure 1	— In object localization tasks, most often the minimal bounding rectangle, i.e. the bounding box, describes where an object is located in an image, here around the person.....	7
Figure 2	— Object detection of a dog and a person, here both objects are localized using a bounding box, the colour represents the different classes these objects are categorized into. ....	9
Figure 3	— Human pose estimation is often conducted on multiple rigid body based model assumption.....	18